

# VU Research Portal

## **Bijkans begrepen? Feitelijk en vermeend begrip van forensische deskundigenrapportages onder rechters, advocaten en deskundigen**

de Keijser, J.W.; Elffers, H.; Kok, R.M.; Sjerps, M.J.

2009

[Link to publication in VU Research Portal](#)

### ***citation for published version (APA)***

de Keijser, J. W., Elffers, H., Kok, R. M., & Sjerps, M. J. (2009). *Bijkans begrepen? Feitelijk en vermeend begrip van forensische deskundigenrapportages onder rechters, advocaten en deskundigen*. Boom Juridische uitgevers.

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### **E-mail address:**

[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)

Bijkans begrepen?



# Bijkans begrepen?

*Feitelijk en vermeend begrip van forensische deskundigen-  
rapportages onder rechters, advocaten en deskundigen*

J.W. de Keijser

H. Elffers

R.M. Kok

M.J. Sjerps

**nsCr**  
Nederlands Studiecentrum  
Criminaliteit en Rechtshandhaving

NEDERLA  
NDSFORE  
NSISCHIN  
STITUUT



Boom Juridische uitgevers  
Den Haag  
2009

Dit project werd financieel mogelijk gemaakt door het Nederlands Forensisch Instituut (NFI).

Omslagontwerp en opmaak binnenwerk: Textcetera, Den Haag

Illustratie omslag: portret van Marie Jean Antoine Nicolas de Caritat, marquis de Condorcet (1743-1794). Hij was de eerste die waarschijnlijkheidsrekening op rechterlijke beslissingen toepaste in zijn *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix* (1785).

© 2009 J.W. de Keijser, H. Elffers, R.M. Kok & M.J. Sjerps / Boom Juridische uitgevers

*Behoudens de in of krachtens de Auteurswet gestelde uitzonderingen mag niets uit deze uitgave worden veeleenvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen of enige andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever.*

Voor zover het maken van reprografische veeleenvoudigingen uit deze uitgave is toegestaan op grond van artikel 16h Auteurswet dient men de daarvoor wettelijk verschuldigde vergoedingen te voldoen aan de Stichting Reprorecht (Postbus 3051, 2130 KB Hoofddorp, [www.reprorecht.nl](http://www.reprorecht.nl)). Voor het overnemen van (een) gedeelte(n) uit deze uitgave in bloemlezingen, readers en andere compilatiewerken (art. 16 Auteurswet 1912) kan men zich wenden tot de Stichting PRO (Stichting Publicatie- en Reproductierechten Organisatie, Postbus 3060, 2130 KB Hoofddorp, [www.cedar.nl/pro](http://www.cedar.nl/pro)).

*No part of this book may be reproduced in any form, by print, photoprint, microfilm or any other means without written permission from the publisher.*

ISBN 978-90-8974-225-4

NUR 824

[www.bju.nl](http://www.bju.nl)

## Dankwoord

Dit onderzoek werd financieel mogelijk gemaakt door het Nederlands Forensisch Instituut (NFI).

Het welslagen van een groot empirisch project als het onderhavige is afhankelijk van de medewerking van vele personen en instanties. Vooraleerst danken wij de deelnemers aan onze experimenten, die bereid waren tijd en energie te steken in het lezen en beantwoorden van lastige vragen over complexe materie. Die deelnemers waren in de voorstudie enkele honderden studenten aan verschillende technische en juridische faculteiten in het land. De hoofdstudie werd mogelijk gemaakt dankzij de deelname van grote aantallen forensisch deskundigen en deskundigen in opleiding van het NFI, rechters en raadsheren, en advocaten.

De medewerkers van de benaderde universiteiten, de Raad voor de rechtspraak, de directie en afdelingshoofden van het NFI, de presidenten en strafsectorvoorzitters van de benaderde gerechten, en het bestuur van de Nederlandse Vereniging voor Strafrechtadvocaten (NVSA), hebben allen een onmisbare rol gespeeld bij het succesvol uitvoeren van het onderzoek.

Ten slotte gaat onze dank en grote waardering uit naar de begeleidingscommissie, bestaande uit voorzitter prof. dr. mr. C.C.J.H. Bijleveld (NSCR, VU), mr. D. Aben (Hof Amsterdam), prof. dr. A.P.A. Broeders (Universiteit Maastricht, The Maastricht Forensic Institute), drs. J. de Koeijer (NFI) en dr. T.B.P.M. Tjin-A-Tsoi (NFI). De diepgravende discussies met en suggesties van de begeleidingscommissie waren van groot belang.



# Inhoud

<b>Samenvatting</b>	9
<b>Introductie</b>	11
Sporen, schuld en de waarde van het bewijs	11
De prosecutor's fallacy	12
De aannemelijkheidsverhouding	13
De 'logisch correcte rapportage'	14
De aannemelijkheidsverhouding en de regel van Bayes	15
Hoe pakt de methode van de aannemelijkheidsverhouding uit in de praktijk?	17
Een korte schets van het onderzoek	17
Vraagstelling hoofdstudie	19
<b>Studie 1: voorwerk</b>	21
Doel en onderzoeksvragen	21
Proefpersonen	22
Materiaal	22
Vragenlijst	23
Onderzoeksopzet	25
Procedure en respons	26
<b>Resultaten Studie 1</b>	27
Feitelijk begrip	27
Vermeend begrip en kennisoverschatting	29
Bèta's versus juristen	30
Conclusies Studie 1; opmaat naar Studie 2	31
<b>Studie 2: voorwerk</b>	33
Inleiding en onderzoeksvragen	33
Meten van feitelijk begrip	34
Materiaal	36
Design	38
Procedure	39
Werving strafrechtadvocaten	39
Werving rechters	39
Werving deskundigen	40
Respons	40



<b>Resultaten Studie 2</b>	43
Criteriumstelling	43
Feitelijk begrip: individuele stellingen	43
Open vragen	46
Gesommeerde scores als meting van feitelijk begrip	48
Feitelijk begrip, per casus en tezamen	49
Geen wezenlijk verschil van feitelijk begrip tussen de twee casus	51
Intermezzo: over feitelijk begrip	51
Vermeend begrip en feitelijk begrip: een kennisoverschatting?	52
Vertrouwdheid met kwantitatief materiaal in het algemeen	54
Wordt een visueel uitgedrukte likelihood ratio beter begrepen?	54
Kwantitatieve voorstelling bij verbale labels aannemelijkheidsverhouding	55
Begrip verklaren	56
Enkele beschouwingen over ‘het probleem’	59
 <b>Conclusies</b>	 63
 <b>Hoe nu verder?</b>	 69
Communicatieprobleem	69
Geen weg terug	69
Wat is de oplossing niet?	70
Studiegroep nodig	70
 <b>Literatuur</b>	 73
 <b>Appendix 1 Onderzoeksmateriaal: deskundigenrapportages</b>	 75
 <b>Appendix 2 De cijfers</b>	 81
 <b>Appendix 3 Gehele vragenlijst, verbale variant, voor juristen</b>	 89
 <b>Over de auteurs</b>	 101

## Samenvatting

Wanneer deskundigen van het Nederlands Forensisch Instituut (NFI) een rapportage opstellen ten behoeve van de rechtbank, wordt de conclusie tegenwoordig in toenemende mate gesteld in termen van aannemelijkheidsverhoudingen (de zogeheten 'logisch correcte methode'). Een voorbeeld naar aanleiding van gezichtsvergelijkend onderzoek, waarbij beelden van een beveiligingscamera worden vergeleken met een foto van de verdachte, luidt: 'De bevindingen van het onderzoek aan de hand van het geselecteerde beeldmateriaal zijn *veel waarschijnlijker* als de daarop afgebeelde persoon dezelfde persoon is als de persoon op de pasfoto, dan als het een andere persoon betreft.'

In deze studie is experimenteel onderzocht of deze nieuwe methodiek begrepen wordt door hen die met zulke rapporten werken, welke fouten er toch worden gemaakt, en of het begrip dat men meent te hebben van de rapportages overeenkomt met de feitelijke mate van begrip. Er is een voorstudie verricht onder studenten. Het hoofdonderzoek is gehouden onder 118 strafrechters, 69 strafrechtadvocaten en 99 medewerkers van het NFI, die een tweetal rapporten te lezen kregen en hun interpretatie daarvan gaven.

Zeven vragen over wat de rapporten zouden verwoorden, werden aan de proefpersonen voorgelegd. De rechters en advocaten wisten het juiste antwoord op gemiddeld 4,3 vragen, maar ook de mensen van het NFI zelf komen gemiddeld niet boven de 5,2 goede antwoorden uit. Fouten als de zogeheten *prosecutor's fallacy* (drogrede van de aanklager) komen veelvuldig voor. De respondenten menen daarentegen zelf in het algemeen de rapporten goed te begrijpen. Er is sprake van een kennisoverschatting: vermeend begrip overstijgt feitelijk begrip. Terwijl ook individuele stellingen overwegend correct geïnterpreteerd werden, constateren wij dat vooral door rechters en advocaten te veel interpretatiefouten worden gemaakt om van goed begrip van de conclusies te kunnen spreken. Die bevinding kwalificeren wij als 'zorgelijk'. Immers, waar rechters de conclusies in deskundigenrapportages nogal eens geheel of gedeeltelijk verkeerd begrijpen – zonder zich daarvan bewust te zijn – kunnen zij het vormen van een overtuiging over het tenlastegelegde niet optimaal vervullen. Dat betekent uiteraard niet dat hierdoor op grote schaal ten onrechte wordt veroordeeld of ten onrechte wordt vrijgesproken. Hoe vaak onbegrip tot andere beslissingen leidt dan die welke zouden zijn genomen wanneer de rechter de rapportage wel correct zou hebben geïnterpreteerd, is onbekend. Dat is hier niet onderzocht. Maar het staat natuurlijk buiten kijf dat de kwaliteit van de strafrechtspleging erbij gebaat is als conclusies worden begrepen zoals deze door de deskundigen zijn bedoeld. *Mutatis mutandis* geldt deze

overweging ook voor andere betrokkenen in de rechtszaal, wier taak het is de rechter voor te lichten (deskundigen), respectievelijk de betekenis van het bewijsmiddel voor hun cliënt naar waarde te schatten (advocaten).

# Introductie

## SPOREN, SCHULD EN DE WAARDE VAN HET BEWIJS

Wanneer er bepaalde sporen op de plaats van een delict worden aangetroffen en er een verdachte in beeld komt, wordt regelmatig aan deskundigen verzocht de rechtbank voor te lichten of de gevonden sporen te rijmen zijn met de veronderstelling dat die sporen door de verdachte zijn achtergelaten.<sup>1</sup> Als voorbeeld: wanneer er op de plaats van het delict een schoenspoor wordt aangetroffen en bij verdachte een schoen in beslag is genomen, wordt aan schoenspoordeskundigen gevraagd zich uit te spreken over de vraag of dat schoenspoor en de gevonden schoen met elkaar te rijmen zijn.

Meestal kan een deskundige in zulke gevallen niet tot een absolute uitspraak komen, in de vorm: 'Dit spoor is wel/niet gemaakt door de schoen van de verdachte', maar moet hij zijn toevlucht nemen tot een waarschijnlijkheidsuitspraak.

Tot voor kort was het gebruikelijk (Broeders, 2005) dat de deskundige in zo'n geval op grond van zijn expertise een uitspraak deed in de trant van:

- [1] Gegeven dat we dit spoor hebben aangetroffen, is het ... dat het schoenspoor door de schoen van verdachte is gemaakt.

Op de puntjes werden dan kwalificaties ingevuld als 'met aan zekerheid grenzende waarschijnlijkheid, hoogstwaarschijnlijk/zeer waarschijnlijk, waarschijnlijk, zeer wel mogelijk, mogelijk'. In de rechtszaal was men, naar het schijnt, niet ongelukkig met deze vorm van rapportage: de professionele deelnemers aan het proces, rechter, Openbaar Ministerie (OM) en verdediging konden daarmee goed uit de voeten, omdat het duidelijk was dat een meer of minder waarschijnlijke identificatie bijdraagt aan de vestiging van de overtuiging bij de rechter dat verdachte betrokken is bij het ten laste gelegde misdrijf, en als zodanig een schakel vormt bij het al of niet vormen van een overtuigend bewijs.

Op deze manier van rapporteren, die neerkomt op het doen van een herkomstuitspraak, is in de loop der tijd echter veel kritiek gekomen, vooral van de zijde van deskundigen zelf (Robertson & Vignaux, 1995; Broeders, 2005).

---

1 Maar ook in rechercheonderzoek spelen dergelijke rapportages met enige regelmaat een rol.

Deze kritiek is van tweeërlei aard. Het eerste bezwaar is dat in het algemeen de analyses van deskundigen zulke uitspraken eigenlijk niet dragen. Het soort waarschijnlijkheidsuitspraken waartoe deskundigen wel kunnen komen op grond van hun analyses, geeft, op de keper beschouwd, veelal aanleiding tot uitspraken van het type:

- [2] Als het de schoen van de verdachte was die het spoor heeft achtergelaten, dan is het ... waarschijnlijk dat het spoor eruit ziet zoals het mij is voorgelegd.

De formuleringen [1] en [2] zijn niet equivalent (Evet, 1995; Broeders, 1999). De op grond van de deskundigenrapportage mogelijke uitspraak [2] is van hypothetische aard: 'Als de schoen van verdachte afkomstig zou zijn, dan is het zo-en-zo waarschijnlijk dat we sporen aantreffen zoals aangetroffen', terwijl de feitelijk gebruikte formulering [1] de andere kant op redeneert: 'Als we sporen aantreffen zoals aangetroffen, dan is het zo-en-zo waarschijnlijk dat ze van de schoen van verdachte afkomstig zijn.' Als de deskundige het tweede [2] zou willen zeggen, omdat hij dat op grond van zijn analyse kan verdedigen, moet hij uiteraard niet het eerste [1] zeggen, en vice versa.

In een concreet geval: als een bepaalde verdachte schoenmaat 41 heeft en een schoenafdruk heeft achtergelaten, is het aantreffen van een schoenaafdruk met maat 41 uitermate waarschijnlijk. Maar als we een schoenaafdruk van maat 41 aantreffen, is het niet als zodanig heel voor de hand liggend dat de verdachte die heeft achtergelaten; er zijn immers zo veel mensen met schoenmaat 41.

## DE PROSECUTOR'S FALLACY

Een uitspraak in de vorm van [1] waarbij, gegeven de waargenomen sporen, een uitspraak wordt gedaan over de waarschijnlijkheid dat die sporen van verdachte afkomstig zijn, wordt, in navolging van Thompson & Schumann (1987) de 'fout van de getransponeerde conditie' genoemd (in het Engels: *transposed conditional*) of ook wel de 'drogreden van de aanklager' (in het Engels: *prosecutor's fallacy*). Deze benaming vindt haar oorsprong daarin dat een aanklager wel graag een uitspraak van de vorm [1] zou willen horen; een feitelijke herkomstuitspraak. Zoals we gezien hebben, was het evenwel tot voor kort gebruikelijk dat deskundigen in de vorm van [1] rapporteerden, dus een correcte benaming van de drogreden van de aanklager zou ook kunnen zijn 'de oude drogreden van de deskundige'. En ook rechters en raadslieden verwarren [1] en [2] nogal eens. Maar goed, de naam *prosecutor's fallacy* is ingeburgerd, dus daar houden we ons aan.

## DE AANNE MELIJKHEIDSVERHOUDING

Het tweede punt van kritiek is dat enkelvoudige waarschijnlijkheidsuitspraken niet zo veelzeggend zijn, of het nu om uitspraken van de vorm [1] of [2] gaat. De uitspraak dat iets in een bepaalde mate waarschijnlijk is, moet altijd in zijn context worden gezien en worden vergeleken met hoe (on)waarschijnlijk het optreden van het bestudeerde zou zijn geweest onder andere omstandigheden. Om terug te keren naar het voorbeeld van de schoen met maat 41, moet dus niet alleen worden gekeken naar de waarschijnlijkheid van een overeenkomst als de schoen van verdachte (met maat 41) de afdruk heeft achtergelaten, maar ook naar de waarschijnlijkheid van overeenkomst met een willekeurige andere schoen. Die waarschijnlijkheid is goed te schatten, omdat ongeveer bekend is hoeveel personen schoenmaat 41 hebben.

Er wordt dan ook bepleit waarschijnlijkheidsrapportages altijd in de vorm van een *vergelijking* van twee mogelijke standen van zaken te verrichten (bijvoorbeeld: Robertson & Vignaux, 1995). Daarbij berekent men niet alleen:

[2] de waarschijnlijkheid van het aantreffen van het spoor, als de afdruk door de schoen van verdachte is achtergelaten,

maar ook:

[2'] de waarschijnlijkheid van het aantreffen van het spoor, als de afdruk door een willekeurige andere schoen is gemaakt

en men vergelijkt beide grootheden, meestal door ze op elkaar te delen. De verhouding van [2] tot [2'] is dan waar het op aankomt. Deze verhouding wordt de *aannemelijkheidsverhouding van het onderzoeksmateriaal onder beide hypothesen* genoemd. Alleen als de waarschijnlijkheid onder [2] groter is dan die onder [2'], dus als de aannemelijkheidsverhouding groter is dan 1, pleit een en ander ten nadele van de verdachte.<sup>2</sup>

2 Merk op dat er uiteraard nogal veel afhangt van de vraag *welke* rivaliserende hypothese men onder [2'] in het oog neemt. Als we niet 'een willekeurige andere schoen' zouden nemen, maar 'een willekeurige andere sportschoen' of 'een willekeurige andere herenschoen' of 'de schoen van medeverdachte X', krijgt men in het algemeen geheel andere uitkomsten. Dat is in de praktijk van het strafrecht uiteraard een zeer wezenlijke zaak, maar in de literatuur over aannemelijkheidsverhoudingen wordt dit probleem tot op zekere hoogte stiefmoederlijk behandeld. In dit rapport laten we deze kwestie ook terzijde. Zie echter: Evett (1995), Sjerps (2000).

Deze aannemelijkheidsverhouding wordt ook wel de *diagnostische waarde* van een specifiek stuk bewijs genoemd.<sup>3</sup> Ook zijn de termen *aannemelijkheids-quotiënt* en *likelihood ratio* gebruikelijk.

Juist aan de hand van de aannemelijkheidsverhouding is te begrijpen waarom bijvoorbeeld een DNA-match zo sterk pleit tegen een verdachte. Immers, de noemer in de aannemelijkheidsverhouding is in zo'n geval extreem klein. De waarschijnlijkheid van een overeenkomst onder de rivaliserende hypothese ('afkomstig van een willekeurig persoon') is bij DNA<sup>4</sup> *kleiner* dan 0,000000001. Als de noemer zo klein wordt, wordt de aannemelijkheidsverhouding heel groot. Bij een matchende bloedgroep daarentegen is de diagnostische waarde veel minder groot, juist omdat in dat geval de waarschijnlijkheid van een overeenkomst aanzienlijk groter is als de rivaliserende hypothese waar is, waardoor dus de verhouding aanzienlijk kleiner is.

### DE 'LOGISCH CORRECTE RAPPORTAGE'

Deze kijk op hoe men waarschijnlijkheidsoordelen behoort te presenteren, is door het NFI omarmd en op grond daarvan heeft men een nieuwe vorm van rapporteren van waarschijnlijkheidsoordelen ingevoerd, die de 'logisch correcte rapportage' wordt genoemd (vgl. Meulenbroek, 2008). Het NFI streeft ernaar niet langer<sup>5</sup> te rapporteren in de vorm van [1], noch in de vorm van [2], maar alleen in de vorm van aannemelijkheidsverhoudingen [2]/[2'].<sup>6</sup>

Wanneer men beschikt over voldoende kennis en informatie die relevant is voor beide hypothesen, zou men een kwantitatief uitgedrukte aannemelijkheidsverhouding kunnen rapporteren. Vaak is de beschikbare informatie daarvoor niet goed genoeg en kan men zich slechts uitspreken over de orde van grootte van de resulterende aannemelijkheidsverhouding. Daarom kiest het NFI er in de meeste gevallen voor om een *verbaal geformuleerde variant* van de aannemelijkheidsverhouding te gebruiken.<sup>7</sup>

3 Zie ook over de diagnostische waarde van bewijs: Crombag, Van Koppen & Wagenaar (2002).

4 Gemakshalve gaan we in dit voorbeeld uit van een volledig profiel.

5 De invoering van deze methodiek is nog niet geheel voltooid. Niet alle deelgebieden hebben de 'logisch correcte vorm' reeds geïmplementeerd, maar daar wordt binnen het NFI wel naar gestreefd.

6 Zie *Vakbijlage: De reeks waarschijnlijkheidstermen van het NFI en het Bayesiaanse model voor interpretatie van bewijs* (NFI, 2008).

7 In de literatuur zijn overigens vele verschillende voorstellen te vinden voor de vertaling van kwantitatieve aannemelijkheidsverhoudingen in woorden, en er is amper evidentie dat verschillende formuleringen tot verschillende interpretaties aanleiding geven (Evet, 1987; Brun & Teigen, 1988; Robertson & Vignaux, 1995; Olson & Budescu, 1997; Sjerps & Biesheuvel, 1999; Broeders, 1999; Champod & Evett, 2000). Murphy et al. (1980) lieten,

Concreet komt dat erop neer dat men uitspraken doet in de vorm van:

- [3] We bekijken twee hypothesen:  $H_1$  is dat de schoen van verdachte het spoor heeft gemaakt;  $H_2$  is dat een willekeurige andere schoen het spoor heeft gemaakt. Op grond van analyse blijkt dat het ... is dat het sporenmateriaal wordt aangetroffen als  $H_1$  waar is, dan dat het sporenmateriaal wordt aangetroffen als  $H_2$  waar is.

Daarbij staat op de plaats van de puntjes een van de termen ‘zeer veel waarschijnlijker’, ‘veel waarschijnlijker’, ‘waarschijnlijker’, ‘iets waarschijnlijker’ of ‘ongeveer even waarschijnlijk’. In het voorkomende geval dat de aannemelijkheidsverhouding kleiner is dan 1 (het sporenmateriaal treedt met grotere waarschijnlijkheid op als  $H_2$  waar is dan als  $H_1$  waar is), rapporteert men ‘omgekeerd’: in uitspraak [3] worden dan de termen  $H_1$  en  $H_2$  omgewisseld.

Deze vorm van rapporteren [3] mag een verbetering worden genoemd ten opzichte van de vroegere manier [1], omdat het twee misverstanden zou kunnen voorkomen: enerzijds de drogreden van de aanklager (daarom wordt [1] door [2] vervangen) en anderzijds het los interpreteren van kansuitspraken (daarom wordt vervolgens [2] door [3] vervangen). Merk op dat wie in de modus [1] rapporteert altijd de aanklagersdrogreden maakt, dus aan dat front kan enkel winst worden geboekt.

## DE AANNEMELIJKHEIDSVERHOUDING EN DE REGEL VAN BAYES

Met de term ‘logisch correcte’ vorm van rapporteren wordt meestal bedoeld dat men de conclusie van het onderzoek in de vorm van een aannemelijkheidsverhouding, dus in de vorm [3], presenteert. Anderen stellen ‘logisch correct rapporteren’ echter gelijk met wat ook wel bayesiaans rapporteren wordt genoemd. Om dat te begrijpen, is het nodig de bayesiaanse evidentieleer kort te bespreken. In die leer (vgl. Kerkmeester, 2005; Sjerps, 2008) wordt ervan uitgegaan dat men de waarde van de ter beschikking staande evidentie over een onzekere zaak kan uitdrukken in waarschijnlijkheidstermen en dat een rechter *voordat* hij het bewijsmateriaal bekijkt zijn heersende onzekerheid of de verdachte schuldig is aan het tenlastegelegde uitdrukt in de vorm van een waarschijnlijkheid dat de verdachte schuldig is, de zogeheten *a priori* waarschijnlijkheid van schuld. Dan bestudeert de rechter het bewijsmateriaal, uitgedrukt in de vorm van de aannemelijkheidsverhouding van het materiaal onder hypothese  $H_1$  (schuld) en  $H_2$  (geen schuld), om op grond daarvan zijn onzekerheid over schuld bij te stellen tot een *a posteriori* waarschijnlijkheid

---

in de context van de interpretatie van weersvoorspellingen, zien dat leken niet zo houden van verbaal geformuleerde waarschijnlijkheidsuitspraken.



van schuld (en op grond daarvan zijn overtuiging van schuld al of niet te vestigen). Op basis van de uiteindelijke *a posteriori* waarschijnlijkheid van schuld besluit de rechter al of niet te veroordelen. In de bayesiaanse leer geldt dus dat de *a posteriori* waarschijnlijkheid een bepaalde wel gedefinieerde functie is van de *a priori* waarschijnlijkheid enerzijds en anderzijds de aannemelijkheidsverhouding. Die functie wordt *de regel van Bayes* genoemd, en luidt:

$$[4] \text{ posterior odds} = \text{prior odds} \times \text{aannemelijkheidsverhouding}^8$$

Die regel is natuurlijk alleen echt toepasbaar wanneer zowel de *prior odds* als de aannemelijkheidsverhouding in kwantitatieve termen zijn uitgedrukt, anders kan men immers niet rekenen. Toch wordt het redeneerschema ook, bij analogie, toegepast als men de aannemelijkheidsverhouding alleen maar in verbale termen heeft weergegeven, zoals bij de logisch correcte methode het geval is (dus in termen van ‘veel waarschijnlijker’, enz.). Dan interpreteert men een verbale vorm van [4] in de zin dat de *posterior odds* in ieder geval veel groter zijn dan de *prior odds*. Voor verschillende onderdelen van het totale bewijs in een zaak kan in beginsel de regel van Bayes herhaaldelijk worden toegepast, waarbij *posterior odds* telkens weer worden bijgesteld door ze als *prior odds* in verband te brengen met het volgende bewijs.<sup>9</sup>

Niet alle rechters zijn echter gecharmeerd van het bayesiaanse model en sommigen willen expliciet niet in termen van *a priori* en *a posteriori* kansen redeneren. Vrij algemeen is men van mening dat dit aan de rechter zelf is, zowel of hij bayesiaans wenst te opereren en al helemaal welke *a priori* waarschijnlijkheid hij wenst aan te nemen. Rechters die zich in dit model kunnen vinden, zullen op grond van [4] hun *a posteriori* waarschijnlijkheid berekenen, en rechters die dat geen aantrekkelijke gedachte vinden, zullen op andere wijze gebruikmaken van de informatie dat het bewijsmiddel al of niet veel of weinig steun geeft aan de ene of de andere hypothese. De deskundige dient zich te beperken tot het rapporteren van de zijns inziens geldende aannemelijkheidsverhouding en het aan de rechter overlaten of en hoe hij dat in een

8 Het is gebruikelijk om de regel van Bayes in termen van ‘odds’ te formuleren. Als  $p_1$  de waarde van de *a priori* waarschijnlijkheid is dat iets het geval is, dan wordt met *odds* ( $p_1$ ) bedoeld de waarde  $p_1/(1-p_1)$ . Anders dan Engelsen hebben veel Nederlanders nogal moeite met het begrip *odds*, zij spreken liever in termen van kansen. Men kan de regel van Bayes ook heel wel formuleren in termen van de *a priori* en *a posteriori* waarschijnlijkheden  $p_1$  en  $p_2$  zelf, in plaats van in termen van *a priori odds* ( $p_1$ ) en *a posteriori odds* ( $p_2$ ), maar dan ziet de regel er wat minder eenvoudig uit.

9 Voor veel auteurs geldt dat zij het rapporteren in termen van een aannemelijkheidsverhouding onverbrekelijk verbonden achten met het bayesiaanse redeneerschema. Anderen menen dat dat niet noodzakelijk is. Ook wie zich daar niet in kan vinden, kan heel wel van mening zijn dat de aannemelijkheidsverhouding een prima methode is om de relatieve steun uit te drukken die een bewijsmiddel aan de ene of de andere hypothese geeft.

bayesiaans beslisschema wenst in te passen. Voor dit rapport zullen we ons daarom concentreren op de interpretatie van de aannemelijkheidsverhouding als zodanig, zonder ons bezig te houden met de grootte van *a priori* kansen.

## HOE PAKT DE METHODE VAN DE AANNEMELIJKHEIDSVERHOUDING UIT IN DE PRAKTIJK?

Hoezeer men ook de argumentatie achter deze vorm van rapporteren kan onderschrijven, het is daarmee nog niet vanzelfsprekend dat de gebruikers van deze rapportage, politie, OM, zittende magistratuur (ZM) en advocatuur, daar ook blij mee zijn en er goed mee kunnen omgaan. Integendeel, niet-systematisch verzamelde indrukken van deskundigen die op deze manier rapporteren, geven aan dat menig procesdeelnemer er niet gelukkig mee zou zijn, het gebruik van twee hypothesen verwarrend en complex zou vinden en dat men terug zou verlangen naar de oude rapportagevorm. Ook de aanklagersdrogreden zou nog veel voorkomen.<sup>10</sup> Het is alleszins aannemelijk dat de indrukken van deskundigen in deze de vinger op de wonde leggen. Ook in het buitenland is immers uit onderzoek gebleken dat rechters (en jury's) grote problemen hebben met het interpreteren van deskundigenrapportages nieuwe stijl (Thompson, 1989; Gatowski et al., 2001). Al met al zien wij aanleiding onderzoek te doen naar de mate waarin rechters, advocaten en deskundigen de nieuwe rapportagemethode begrijpen en accepteren. Het onderhavige rapport doet daar verslag van.

## EEN KORTE SCHETS VAN HET ONDERZOEK

In de onderhavige studie zijn deskundigenrapporten in de vorm van waarschijnlijkheidsrapportages 'nieuwe stijl' voorgelegd aan een aantal betrokkenen bij het strafproces (rechters, advocaten, deskundigen). Er is nagegaan of deze rapporten worden begrepen, welke fouten er worden gemaakt en hoe men tegen deze vorm van rapporteren aankijkt. Om dat goed te kunnen doen, is het uiteraard noodzakelijk precies te zijn in wat dan wel de correcte interpretatie van een rapport is. In onze studie maken wij onderscheid tussen *feitelijk begrip* en *vermeend begrip*. Wij achten dit onderscheid van groot belang, omdat wij en anderen regelmatig professionele procesdeelnemers hebben ontmoet die in volle overtuiging van hun eigen goede begrip van een deskundigenrapport in even volle overtuiging een *prosecutor's fallacy* als de juiste interpretatie van de conclusie exclameerden.

<sup>10</sup> Daarentegen lieten Murphy et al. (1980) zien dat het Amerikaanse publiek bij het verwoorden van weersvoorspellingen zelf in meerderheid gebruikmaakt van aannemelijkheidsverhoudingen.

*Feitelijk begrip* definiëren wij daarom als begrip van de conclusie, zoals deze daadwerkelijk door de deskundige en in de hiervoor beschreven zin, is bedoeld. Daarentegen is *vermeend begrip* de mate waarin men zelf die conclusie meent te begrijpen. In een ideale wereld is vermeend begrip hetzelfde als feitelijk begrip. Is dat niet het geval, in een minder ideale wereld, als vermeend begrip groter is dan feitelijk begrip, dan spreken wij van een *kennisoverschatting*. In het dagelijks leven hoeft een overschatting van eigen kennis niet vaak een erg groot probleem te zijn. Dat levert vooral voor degene die zichzelf overschat zo nu en dan een onaangename verrassing op, zoals een volstrekt onverwachte dikke onvoldoende voor een wiskundetoets. In de context van het strafrecht mogen we hogere eisen stellen aan de relatie tussen vermeend begrip en feitelijk begrip bij professioneel betrokkenen. Binnen die context worden immers, op basis van vermeend begrip, beslissingen genomen die ingrijpend zijn voor het leven van anderen. Wij mogen verwachten dat rechters de bouwstenen van hun beslissingen feitelijk zo goed begrijpen als zij denken ze te begrijpen.

Om na te gaan of onze respondenten een feitelijk juist begrip van de rapporten vertonen, legden we hen een aantal stellingen voor over de interpretatie van de rapporten, sommige correct, andere incorrect, en gingen we na hoeveel en welke stellingen terecht of niet als fout of juist werden geclassificeerd. Tot de foute stellingen behoren enkele varianten op de *prosecutor's fallacy*. Ook werd een variant van de zogeheten *defense fallacy* (drogreden van de verdediging) voorgelegd. Dat is een door Thompson & Schumann (1987) zo gedoopte misvatting, die vaak als volgt wordt verwoord: als een bewijsmiddel (bijvoorbeeld een bloedgroep zoals aangetroffen in bloed op de plaats van een delict, niet zijnde van het slachtoffer) in een zekere deelgroep van omvang A van alle mensen voorkomt, en het komt ook bij de verdachte voor, dan is de kans dat juist de verdachte dat spoor heeft achtergelaten  $1/A$ . Als zodanig is het eigenlijk een variant op de *prosecutor's fallacy*, daar het immers een uitspraak over de waarschijnlijkheid van een hypothese op grond van het bewijs ten onrechte afleidt uit gegevens die informatie geven over de waarschijnlijkheid van bewijs op grond van een hypothese. Als A groot is, klinkt zo'n redenering, indien niet doorzien, vaak voordelig voor de verdachte, vandaar de naam *defense fallacy*.

Het onderzoek kent twee fasen. De eerste fase, de *voorstudie* of Studie 1, was bedoeld om het instrumentarium voor de 'meting van begrip' te construeren en uit te testen en maakt gebruik van studenten als proefpersonen. De tweede fase, de *hoofdstudie* of Studie 2, dient om na te gaan hoe het begrip feitelijk en vermeend blijkt te liggen onder de werkelijk betrokkenen, te weten strafrechters, strafrechtadvocaten en medewerkers van het NFI die

waarschijnlijkheidsrapportages opstellen.<sup>11</sup> In de hoofdstudie is men tevens gevraagd te reageren op enkele algemeen beschouwende vragen over rapporteren in de vorm van aannemelijkheidsverhoudingen.

## VRAAGSTELLING HOOFDSTUDIE

Worden deskundigenrapportages waarin de conclusie geformuleerd is in termen van een verbale verwoording van de aannemelijkheidsverhouding van de gevonden sporen ten opzichte van twee rivaliserende hypotheses, correct geïnterpreteerd door strafrechters, door strafrechtadvocaten en door deskundigen die zulke rapportages opstellen?<sup>12</sup>

- (1) Hoe schatten zij hun eigen begrip in (vermeend begrip)?
- (2) Welke feitelijke fouten (incorrecte interpretaties) worden gemaakt (feitelijk begrip), en in welke mate?
- (3) In hoeverre is er sprake van een kennisoverschatting?
- (4) Welke kenmerken van de beoordelaars hangen samen met het vóórkomen van fouten (ervaring, het gevolgd hebben van relevante cursussen, vertrouwdheid met cijfers<sup>13</sup>)?
- (5) Welke kenmerken van de gerapporteerde aannemelijkheidsverhouding (groot of klein, ten gunste van de ene of de andere hypothese) hangen samen met begrip?
- (6) Hoe kijken de procesdeelnemers aan tegen het werken met deze rapportagevorm?
- (7) Zijn er aanpassingen in de manier van rapporteren denkbaar die het begrip ten goede komen?

11 Om reden van tijd en budget moesten wij helaas afzien van het bepalen van begrip bij politiemensen en leden van het Openbaar Ministerie.

12 In de voorstudie zijn enkele van deze vragen en prealabele vragen bestudeerd en beantwoord, en dat heeft aanleiding gegeven om in de hoofdstudie sommige vragen meer nadruk te geven dan andere.

13 Vergelijk Kaasa et al. (2007), die suggereren dat er een wezenlijk onderscheid in begrip is tussen mensen die van zichzelf zeggen dat ze goed dan wel slecht kunnen omgaan met cijfers.



# Studie 1: voorwerk

## DOEL EN ONDERZOEKSVRAGEN

In de hoofdstudie wordt het begrip door strafrechters, strafrechtadvocaten en deskundigen van verbale *likelihood ratio's* in deskundigenrapporten onderzocht. Er is een voorstudie uitgevoerd alvorens de professionele procesdeelnemers te benaderen en te ondervragen. Belangrijke reden daarvoor was dat de beoogde populaties van professionele procesdeelnemers te klein zijn om het op korte termijn nog eens over te doen wanneer er technische, methodologische of theoretische problemen zouden optreden bij of na afloop van de meting. Studie 1 diende daarom enerzijds als 'speeltuin' om verschillende manieren van bevraging over begrip van deskundigenrapportages uit te proberen, anderzijds om reeds enig inhoudelijk licht op enkele begripgerelateerde zaken te laten schijnen.

Als gezegd, maken wij onderscheid tussen feitelijk begrip en vermeend begrip. *Feitelijk begrip* is begrip van de conclusie, zoals deze daadwerkelijk door de deskundige is bedoeld. *Vermeend begrip* is de mate waarin de respondent de conclusie meent te begrijpen. We mogen hopen dat bij professionele procesdeelnemers feitelijk begrip en vermeend begrip overeenkomen. Is dat niet het geval, en met name als vermeend begrip groter is dan feitelijk begrip, dan spreken wij van een *kennisoverschatting*. Hoe groot een eventuele kennisoverschatting is, hopen wij in kaart te brengen.

De volgende vier concrete vragen waren leidend bij de opzet van de voorstudie:

1. Is feitelijk begrip te meten?
2. Is er verschil tussen vermeend begrip bij ondervraagden en feitelijk begrip; is er sprake van een kennisoverschatting?
3. Zijn mensen met gedegen wiskundige kennis beter in staat verbale *likelihood ratio's* correct te interpreteren dan mensen die die achtergrond missen?
4. Is hoogte (groot of klein) of vorm (ten gunste van de ene of andere hypothese) van de *likelihood ratio* van invloed op begrip?

Terwijl het ontwikkelen van het meetinstrument, het verkennen van een eventuele kennisoverschatting, en het variëren van de *likelihood ratio* direct ten dienste stonden van de hoofdstudie, lag een eigenstandig doel van deze Studie 1 in beantwoording van de vraag of mensen met een gedegen wiskun-

dige achtergrond technisch forensische conclusies in de vorm van *likelihood ratio's* beter begrijpen dan juristen. Indien die vraag namelijk bevestigend kan worden beantwoord, zou dat van praktische betekenis kunnen zijn voor eventuele aanbevelingen voor de opleiding van juridisch geschoolde professionele procesdeelnemers die met dergelijke rapportages moeten werken.

## PROEFPERSONEN

Voor deelname aan deze studie zijn studenten benaderd. Om vraag 3, naar de invloed van wiskundige kennis op begrip van verbale *likelihood ratio's*, te kunnen onderzoeken, is enerzijds specifiek geworven onder *bètastudenten*, van wie een degelijke kennis van wiskunde en statistiek mag worden verwacht, anderzijds specifiek onder *rechtenstudenten*. Bij die laatste groep mag men er redelijkerwijs van uitgaan dat weinigen een stevige kwantitatieve achtergrond hebben.

## MATERIAAL

Omdat vraag 4 in Studie 1 betrekking heeft op variatie in hoogte en aard van de *likelihood ratio*, zijn drie fictieve doch realistische deskundigenrapporten geconstrueerd. In het conclusiegedeelte van twee van die rapporten varieerde de mate waarin de bevindingen waarschijnlijker waren onder hypothese 1 dan onder hypothese 2: *iets waarschijnlijker* respectievelijk *veel waarschijnlijker*. In het derde rapport waren de resultaten *veel waarschijnlijker* onder hypothese 2 dan onder hypothese 1; in dat geval was de *likelihood ratio* dus 'omgeklapt'.

In de toelichting op het onderzoeksontwerp hieronder is te lezen hoe en waarom ervoor gekozen is om aan iedere respondent drie rapporten voor te leggen. Deze keuze noodzaakt ertoe dat de rapporten geheel verschillende onderzoeken dienen te betreffen. Als namelijk alleen het conclusiegedeelte tussen de drie rapporten zou variëren, zou het de deelnemers aan het onderzoek onmiddellijk duidelijk zijn wat de focus van de studie is. Antwoorden op basis van een eerder rapport zouden dan beantwoording van vragen betreffende een volgend rapport zonder meer beïnvloeden. De drie fictieve deskundigenrapporten in Studie 1 betreffen:

- A. Een vergelijkend onderzoek naar de lijmlagen van stukken tape gebruikt bij een overval op een tankstation en tape in beslag genomen bij een verdachte.
- B. Een vergelijkend onderzoek naar beeldmateriaal van een beveiligingscamera van een pinautomaat op straat waar iemand beroofd was en foto's met daarop afbeelding van een verdachte.
- C. Een vergelijkend onderzoek naar een schoenafdruk gevonden op de plaats van een bedrijfsinbraak en een afdruk van de schoen van een verdachte.

Twee van deze rapporten (A en B) staan integraal weergegeven in appendix 1.<sup>14</sup> Het rapport over de schoenafdrukken geven we hier niet integraal weer; wij volstaan met de melding dat dat rapport qua omvang, structuur en opbouw vergelijkbaar is met de andere rapporten. De fragmenten van alle drie de rapporten (A, B en C) die de *conclusies* betreffen, staan hierna weergegeven in box 1.

## Vragenlijst

Na ieder rapport volgden enkele schriftelijke vragen over begrip van de conclusie in dat rapport. Vermeend begrip werd direct gemeten aan de hand van de volgende vraag: ‘Hoe goed of slecht begrijpt u de conclusie van dit rapport?’ Het antwoord kon worden gegeven op een schaal die liep van 1 (‘Ik begrijp er niets van’) tot 7 (‘Ik begrijp het volledig’).

Om feitelijk begrip te meten, formuleerden wij acht stellingen en vroegen aan de respondenten of dit is wat de deskundige in zijn conclusie bedoelt. De helft van de stellingen was een correcte weergave van de conclusie, de andere helft was fout. De fouten betroffen vooral formuleringen van de beruchte valkuilen *prosecutor’s fallacy* en *defense fallacy*. De formulering van de stellingen geschiedde op basis van literatuuronderzoek,<sup>15</sup> eigen ervaring en discussie onderling en met specialisten uit de wetenschap, de rechtspraak en van het NFI.<sup>16</sup> Een voorbeeld van een goede stelling is: ‘De stukken tape kunnen ook van een andere rol tape afkomstig zijn dan van de rol van de verdachte.’ Een voorbeeld van een foute stelling is: ‘Er is iets meer dan 50% kans dat de tape afkomstig is van de rol van de verdachte.’<sup>17</sup> De respondent kon voor iedere stelling op de vraag of dit is wat de deskundige bedoelt, antwoorden met ‘ja’, ‘nee’, of ‘weet niet’. Aan het eind van de vragenlijst werden nog enkele achtergrondvragen aan de deelnemers gesteld.

14 Feitelijk zijn dit de rapporten zoals gebruikt in Studie 2, waarin enkele kleine wijzigingen ten opzichte van Studie 1 waren aangebracht.

15 Voor de formulering van de vragen is afgezien van systematisch variëren over Evett’s hiërarchie van proposities, die de uitspraken in een rapportage onderverdeelt in uitspraken over de bron van het spoor, over of het spoor gerelateerd is aan het delict, of er sprake is van een delict, of de verdachte de dader is, of de verdachte schuldig is, of verdachte strafbaar is.

16 Deze terreinen waren goed vertegenwoordigd bij de leden van de begeleidingscommissie.

17 Het voert te ver om hier alle stellingen te behandelen. Naar aanleiding van ervaringen met deze stellingen in Studie 1, kwam het definitieve blok stellingen tot stand als gebruikt in Studie 2. In de desbetreffende paragraaf hierna wordt dat blok stellingen wel integraal behandeld in de lopende tekst. Zie appendix 2 voor meer details over de individuele stellingen gebruikt in Studie 1.



## Box 1 Conclusies uit de drie fictieve deskundigenrapporten

### Rapport A – Overval tankstation (lijmlagen tape)

De vraagstelling is geïnterpreteerd als het verzoek om een vergelijkend onderzoek te verrichten en na te gaan of de daarbij verkregen resultaten (de bevindingen) beter passen bij de hieronder weergegeven hypothese H1 of bij de (alternatieve) hypothese H2.

*Hypothese 1:*

*De tape waarmee slachtoffer gekneveld was is afkomstig van de rol tape uit de woning van verdachte.*

*Hypothese 2:*

*De tape waarmee slachtoffer gekneveld was is afkomstig van een willekeurige andere rol tape.*

(...)

De bevindingen van het vergelijkend onderzoek zijn *iets waarschijnlijker* als de tape waarmee slachtoffer gekneveld was afkomstig is van de rol tape uit de woning van verdachte (hypothese 1) dan als de tape waarmee slachtoffer gekneveld was afkomstig is van een willekeurige andere rol tape (hypothese 2).

### Rapport B – Beroving op straat (beeldmateriaal)

Op basis van vraagstelling eerste screening heb ik een gezichtsvergelijkend onderzoek verricht en gezien of de resultaten daarvan beter passen bij de hieronder weergegeven hypothese H1 of bij de (alternatieve) hypothese H2.

De hypothesen luiden:

*Hypothese 1:*

*De dader van de beroving zichtbaar op SX1\_C1 is dezelfde persoon als verdachte afgebeeld op foto SX1\_B1.*

*Hypothese 2:*

*De dader van de beroving zichtbaar op SX1\_C1 is niet dezelfde persoon als verdachte afgebeeld op foto SX1\_B1.*

(...)

De bevindingen van de hier gerapporteerde gezichtsvergelijking aan de hand van het geselecteerde beeldmateriaal zijn *veel waarschijnlijker* als de daarop afgebeelde persoon dezelfde persoon is (hypothese 1) dan als het een andere persoon betreft (hypothese 2).

**Rapport C – Bedrijfsinbraak (schoenafdruk)**

Deze vraagstelling is vertaald in de opdracht een vergelijkend schoenspooronderzoek te verrichten en te onderzoeken of de bevindingen daarvan beter passen bij de hieronder weergegeven hypothese H1 of bij de (alternatieve) hypothese H2. Ik formuleer nu de volgende hypothesen voor dit onderzoek:

*Hypothese 1:*

*De rechterschoen van de verdachte maakte het spoor.*

*Hypothese 2:*

*Een willekeurige andere rechterschoen maakte het spoor.*

(...)

Ik concludeer dan ook met betrekking tot mijn eerder geformuleerde hypothesen dat:

‘De bevindingen van het hier gerapporteerde schoenspooronderzoek *veel waarschijnlijker* zijn als een willekeurige andere rechterschoen het spoor maakte (hypothese 2) dan als de rechterschoen van verdachte het spoor maakte (hypothese 1).’

**ONDERZOEKSOPZET**

De onderzoeksopzet (*design*) was zodanig dat iedere deelnemer alle drie de deskundigenrapporten met bijbehorende vragen kreeg voorgelegd. Bij een dergelijke opzet met herhaalde meting bij dezelfde personen (dit wordt wel *within subjects design* genoemd) treedt het gevaar op van volgorde-effecten op de beantwoording. Daar is een simpele oplossing voor: het systematisch variëren van die volgorde, waardoor ieder rapport even vaak als eerste, als tweede en als derde wordt aangeboden. Dat resulteert in drie versies van de totale vragenlijst, die te zien zijn in de rijen van tabel 1.<sup>18</sup>

<sup>18</sup> In de literatuur over opzet van experimenten wordt een dergelijk design een ‘Latijns vierkant’ genoemd.

Tabel 1 Onderzoeksofzet Studie 1

	►	A. Tank	B. Straat	C. Bedrijf
rechtenstudenten				
	►	C. Bedrijf	A. Tank	B. Straat
bètastudenten				
	►	B. Straat	C. Bedrijf	A. Tank

## PROCEDURE EN RESPONS

In februari 2008 werden de gegevens verzameld bij studenten aan de Universiteit Leiden, Universiteit Utrecht, Technische Universiteit Eindhoven en de Vrije Universiteit Amsterdam. De procedure was als volgt. Tegen het eind van een hoorcollege werden wij door de desbetreffende docent kort geïntroduceerd en verzochten wij de studenten te blijven zitten om een vragenlijst in te vullen.<sup>19</sup> De aard van het onderzoek werd daarbij slechts in oppervlakkige bewoordingen medegedeeld. Onder de deelnemers werden VVV-bonnen verloot als extra stimulans om mee te doen. De bètastudenten zijn geworven bij hoorcolleges van de studies Informatica, Wiskunde en Biofarmaceutische Wetenschappen, de rechtenstudenten bij Strafrecht en Criminologie.

De drie versies van de vragenlijsten (zie tabel 1) werden in willekeurige volgorde uitgedeeld aan de studenten. Gemiddeld had men 25 minuten nodig om het materiaal te lezen en de vragen te beantwoorden. Ingevulde vragenlijsten werden ter plekke ingenomen.

De respons was zonder meer goed te noemen en varieerde tussen 52% en 100%; gemiddeld 85%. In totaal hebben 336 studenten aan het onderzoek meegedaan, waarvan 49% rechtenstudenten en 51% bètastudenten.

<sup>19</sup> Wij volgden deze procedure bij negen verschillende hoorcolleges. Dit was niet mogelijk geweest zonder de uitermate sympathieke en behulpzame instelling van de docenten die daar meestal zelfs wat college tijd aan opofferden. Onze dank gaat uit naar prof. dr. mr. M. Moerings, prof. mr. J.A.E. Vervaele, dr. J.R. Blad, dr. R. van der Heijden, dr. M. de Jeu, dr. M. van Eck, dr. H.J. Haverkort, dr. M.A. Jonker en prof. dr. S.D. Swierstra.

# Resultaten Studie 1<sup>20</sup>

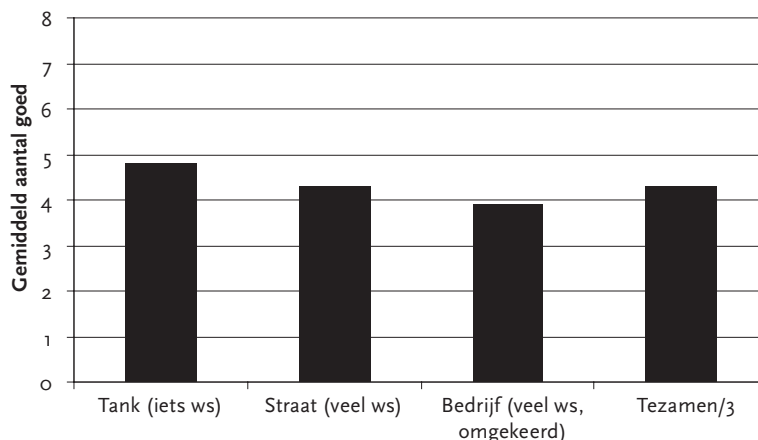
## FEITELIJK BEGRIIP

Feitelijk begrip was, zoals hiervoor beschreven, gemeten door de respondenten een achttal stellingen te laten waarderen. Geeft een stelling volgens de respondent weer wat de deskundige beoogt te zeggen? Het antwoord 'weet niet' is door ons fout gerekend: dat duidt er immers op dat men de conclusie niet begreep.<sup>21</sup> Bij perfect begrip zou de maximale score per casus een '8' zijn; dat wil zeggen alle acht stellingen bij dat rapport zijn correct geclassificeerd. In figuur 1 is voor ieder van de drie rapporten weergegeven hoeveel stellingen gemiddeld correct werden beantwoord. In de figuur is ook het gemiddelde over drie casus heen weergegeven. Het is te verwachten dat per casus ongeveer de helft van de stellingen correct geclassificeerd wordt door de respondenten als de antwoorden uit puur willekeurig of gokgedrag zouden bestaan. Dat zou betekenen dat bij volstrekte afwezigheid van feitelijk begrip, je op basis van toeval nog wel vier stellingen goed hebt.

Figuur 1 laat zien dat de prestaties van de studenten bijzonder dicht bij dat kansniveau liggen. Bij de beroving van het tankstation en de beroving op straat liggen de gemiddelden met respectievelijk 4,8 en 4,3 weliswaar statistisch significant boven kansniveau, de mate waarin maakt duidelijk dat het erg slecht gesteld is met feitelijk begrip. Gemiddeld over de drie casus heen, ontstijgt het feitelijk begrip de criteriumscore van 4 met niet meer dan drie tiende. Men kan erover twisten hoeveel vragen men goed dient te scoren voor we zouden willen stellen dat de respondent een redelijk begrip van het rapport heeft. Dat dat veel hoger moet zijn dan het niveau van het opgooien van een munt (vier goed) lijkt onbetwistbaar. Dat is hier niet of nauwelijks het geval: de studenten begrijpen de rapporten niet goed.

<sup>20</sup> De presentatie en bespreking van resultaten in de hoofdtypekst geschiedt vooral aan de hand van figuren. In appendix 2 zijn de relevante cijfers terug te vinden, alsmede statistische toetsen.

<sup>21</sup> Op het onderscheid tussen de twee verschillende typen studenten (bèta's en juristen) gaan we hierna pas in. Dat onderscheid, zo zal blijken, is voor de algemene beschrijving van begrip van de conclusies in de drie casus niet relevant.



**Figuur 1** Feitelijk begrip, Studie 1: gemiddeld aantal correct geclassificeerde stellingen ( $N_{\text{tank}} = 324$ ,  $N_{\text{straat}} = 330$ ,  $N_{\text{bedrijf}} = 320$ )

De stellingen waar consequent in alle drie de rapporten de meeste fouten op werden gemaakt, verwoordden op de een of andere wijze de *prosecutor's fallacy* (of *transposed conditional*). Zo bevestigde 69% van de respondenten onterecht over het onderzoek naar de tape (tankstation) dat de conclusie betekent: 'Er is iets meer dan 50% kans dat de stukken tape afkomstig zijn van de rol van de verdachte.' En, bijvoorbeeld, bij de beroving op straat, bevestigde maar liefst 89% dat het veel waarschijnlijker is dat de verdachte de persoon op de camera-beelden is dan dat het iemand anders is die op de camerabeelden staat.

Over verschillen tussen de drie rapporten kan het volgende worden gconstateerd. Er is betrekkelijk weinig verschil in begrip van de rapporten die als conclusie 'veel waarschijnlijker' hebben. Terwijl in het rapport over de bedrijfsinbraak in de vorm van een 'omgekeerde' *likelihood ratio* werd geconcludeerd, dus veel waarschijnlijker onder  $H_2$  dan onder  $H_1$ , maakte dit voor het feitelijk begrip bij onze respondenten kennelijk niet uit. Er is wel enig, doch een klein, verschil tussen het 'tank-rapport' met 'iets waarschijnlijker' en het 'straat-rapport' met 'veel waarschijnlijker'. De *prosecutor's fallacy* wordt in de krachtigere conclusie wat vaker gemaakt dan in de minder krachtige conclusie: respectievelijk 89% (straat, veel waarschijnlijker) tegenover 85% (tank, iets waarschijnlijker).

Ten slotte, er is sprake van zeer geringe correlaties tussen de antwoorden op de individuele begrip-stellingen. Wie een fout antwoord geeft op de ene stelling doet dus niet met veel grotere kans iets fout op een andere stelling. Dat betekent dat het soort fouten dat wordt gemaakt niet veel met elkaar te maken heeft.

## VERMEEND BEGRIP EN KENNISOVERSCHATTING

Vermeend begrip was direct gemeten door de respondent zichzelf te laten plaatsen op een schaal van 1 ('ik begrijp er niets van') tot een maximum van 7 ('ik begrijp het helemaal'). Deze vraag werd voor ieder van de drie conclusies afzonderlijk gesteld voorafgaand aan het klasseren van de stellingen. Figuur 2 laat zien dat onze respondenten in het algemeen menen de conclusies van de deskundigen goed te begrijpen. De conclusie van het rapport over de bedrijfsinbraak mag zich verheugen in het grootste vermeend begrip met een gemiddelde van meer dan 6. Het vermeend begrip onder de studenten is zo groot, dat de scores 6 en 7 tezamen gegeven worden door driekwart van de studenten.



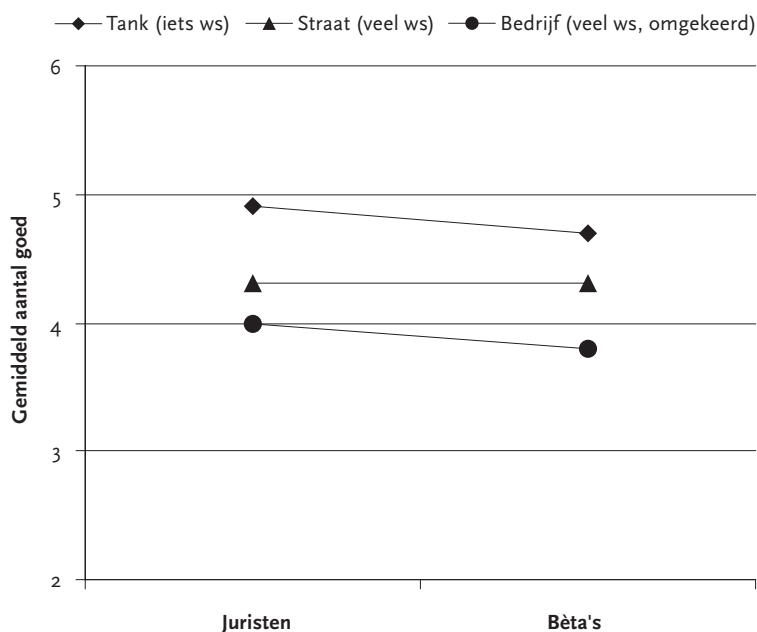
*Figuur 2 Vermeend begrip, Studie 1: gemiddelde score op schaal van vermeend begrip ( $N_{\text{tank}} = 333$ ,  $N_{\text{straat}} = 336$ ,  $N_{\text{bedrijf}} = 334$ )*

In de vorige paragraaf (onder 'Feitelijk begrip') is reeds gebleken hoe slecht de conclusies in de drie rapporten feitelijk worden begrepen. Toch menen dezelfde personen in groten getale het erg goed te hebben begrepen. Kennelijk is er sprake van een grote kennisoverschatting bij de studenten die deelnamen aan ons onderzoek. Men meent de conclusies in de rapporten goed te hebben begrepen, terwijl het feitelijk begrip bol staat van misverstanden. Tussen feitelijk begrip en vermeend begrip bestaat dan ook geen noemenswaardige samenhang.<sup>22</sup>

<sup>22</sup> Correlatiecoëfficiënten tussen vermeend begrip en feitelijk begrip zijn voor Tank: 0; Straat: 0,2; Bedrijf: 0.

## BÈTA'S VERSUS JURISTEN

Wellicht, zo was de achtergrond van onze derde onderzoeksvraag, zijn bèta-studenten veel beter geëquipeerd dan juristen om conclusies in de vorm van een verbaal aanemelijkheidsquotiënt correct te interpreteren. In figuur 3 is te zien of dit het geval is. De figuur illustreert gemiddeld feitelijk begrip (aantal goed) van de drie conclusies voor de juristen vergeleken met de bèta's.<sup>23</sup>



Figuur 3 Feitelijk begrip, Studie 1: juristen vergeleken met bèta's

Figuur 3 laat zien dat een stevig been in de wiskunde geen soelaas biedt voor feitelijk begrip van 'logisch correct' geformuleerde conclusies in deskundigenrapporten. Voor zover er al enig verschil in feitelijk begrip te constateren valt tussen de bètastudenten enerzijds en de rechtenstudenten en criminologen anderzijds, is dat van verwaarloosbare orde van grootte en bovendien in de andere richting dan men zou verwachten.

<sup>23</sup> Onder de term juristen vatten wij hier nu ook de criminologen. Uit de analyses bleek dat tussen deze groepen studenten geen verschillen van betekenis bestaan.

## CONCLUSIES STUDIE 1; OPMAAT NAAR STUDIE 2

De eerste onderzoeksvraag voor deze Studie 1 was of feitelijk begrip zinvol te meten is. Wij hebben feitelijk begrip geoperationaliseerd aan de hand van een blok van acht stellingen die de respondent diende te classificeren als correcte en incorrecte interpretaties van de conclusie van de deskundige. Tussen antwoorden op die stellingen konden wij geen noemenswaardige samenhang constateren. Dat is jammer, omdat de uitkomstmaat die wij hanteerden voor feitelijk begrip een optelsom was van het aantal goed geclassificeerde stellingen. Doorgaans is het wenselijk dat afzonderlijke stellingen binnen zo'n optelling ongeveer hetzelfde meten en dus in redelijke mate samenhangen. Onze verklaring voor dat gebrek aan samenhang is tweeledig. Enerzijds snijden de verschillende stellingen zeer verschillende typen denkfouten aan. Anderzijds zou gesteld kunnen worden dat bij afwezigheid van feitelijk begrip het aannemelijker is dat bepaalde denkfouten wel worden gemaakt dan dat ze niet worden gemaakt. Met name de *prosecutor's fallacy* is voor de meeste mensen erg moeilijk 'te weerstaan'. Daarmee dienen we voorlopig te concluderen dat we zeker niet kunnen volstaan met het meten van feitelijk begrip aan de hand van één enkele of wellicht een twee- of drietal stellingen. Maar het zou toch ook goed zijn enige indicatie te verkrijgen of respondenten wel serieus aandacht hebben besteed aan beantwoording van deze stellingen. Daarbij kan gedacht worden aan een criteriumstelling waarvan met enige zekerheid te voorspellen valt hoe iemand die de vragen serieus bekijkt, zal antwoorden. Bij de bespreking van opzet en vragen voor Studie 2 komen wij hierop terug.

Studie 1 liet zien dat er een forse kennisoverschatting is bij de studenten. Hun vermeend begrip was groot en stond in geen verhouding tot feitelijk begrip. Het is vooralsnog onduidelijk of van een kennisoverschatting in deze mate bij de professionele procesdeelnemers sprake is. Dat is een empirische vraag voor Studie 2.

Tussen de twee groepen studenten die wij onderscheiden in deze studie werd geen verschil in feitelijk begrip (noch in vermeend begrip) aangetoond. Dat was enigszins onverwacht. Deze bevinding geeft te denken over de te verwachten effecten van een eventuele meer exacte component in de opleiding van praktijkjuristen. Anders gesteld: als een wiskundestudent het al niet beter begrijpt dan een jurist, wat zou een algemene cursus wiskunde of statistiek dan kunnen helpen voor praktijkjuristen in de rechtszaal? Maar of een cursus die specifiek is toegespitst op de betekenis van onderhavige waarschijnlijkheidsoordelen enig effect sorteert, is op basis van onze resultaten tot nog toe natuurlijk niet te zeggen.

Ten slotte was de vraag of de hoogte van de aannemelijkheidsverhouding of de vorm waarin die verhouding werd gepresenteerd van invloed is op feitelijk begrip. Het bleek voor de respondenten niet uit te maken of bevindingen



veel waarschijnlijker waren onder  $H_1$  dan onder  $H_2$ , of dat de bevindingen veel waarschijnlijker waren onder  $H_2$  dan onder  $H_1$ . Op basis hiervan achten wij soortgelijke variatie in *vorm* van de *likelihood ratio* niet langer relevant voor verdere studie. De *hoogte* van het aannemelijkheidsquotiënt leek wel van enige invloed op het begrip. Juist de meest beruchte denkfouten als de *prosecutor's fallacy* worden extra verleidelijk bij een grotere *likelihood ratio*. Dit is geheel in lijn met ervaringen met deze materie in de praktijk.

## Studie 2: voorwerk

### INLEIDING EN ONDERZOEKSVRAGEN

De ervaringen met Studie 1 gaven geen aanleiding om een drastisch andere koers te gaan varen bij het onderzoek onder de professionele procesdeelnemers. De aanpak, het materiaal en de onderzoeksprocedure van Studie 2 zijn dan ook vergelijkbaar.

Mede vanwege de lage samenhang tussen de antwoorden op de stellingen in Studie 1, zijn alle stellingen nog eens nauwkeurig heroverwogen. Dit resulteerde in een iets gewijzigd blok stellingen met correcte en foute verwoordingen van wat de deskundige beoogt te zeggen. Daarover meer in de volgende paragraaf.

De hoofdvraag van Studie 2 is hoe goed verschillende professionele procesdeelnemers conclusies van deskundigen, gegoten in de vorm van een verbale *likelihood ratio*, begrijpen. Daartoe hebben wij ons beperkt tot drie professies: strafrechters, strafrechtadvocaten en medewerkers van het Nederlands Forensisch Instituut. De reden dat het Openbaar Ministerie en politie niet betrokken zijn in dit onderzoek, was van puur praktische aard (i.c. beperkte tijd en middelen). Dit onderzoek kan zonder meer gerepliceerd worden onder officieren van justitie en rechercheurs.

In verband met de hoofdvraag is ook getracht een eventuele kennisoverschatting bij deze groepen in kaart te brengen door aandacht te besteden aan vermeend begrip naast feitelijk begrip. Ook is aan de deelnemers een beperkt aantal vragen gesteld over achtergrondkenmerken en opleiding en ervaring, om dat in verband te kunnen brengen met begrip. Uit Studie 1 bleek reeds dat het bij studenten slecht gesteld is met begrip van verbale *likelihood ratio's*, zoals die geformuleerd worden in deskundigenrapportages. Bovendien bleek uit die studie dat wiskundig onderlegde personen (bètastudenten) niet beter presteren dan juristen. Op grond van onze indruk, gebaseerd op reacties uit de professionele rechtspraktijk op verbale *likelihood ratio's* in deskundigenrapporten, vermoedden wij op voorhand dat Studie 2 ook bij juridische professionals zoals rechters en advocaten, veel onbegrip aan het licht zou brengen. Wij besloten daarom Studie 2 niet alleen te gebruiken voor het meten van begrip van verbale *likelihood ratio's*, maar tevens als proeftuin voor een vorm van concluderen die wellicht beter begrepen wordt door niet-deskundigen. Daartoe is

naast de verbale vorm een visuele variant van een aannemelijkheidsverhouding ontwikkeld,<sup>24</sup> waarover hierna meer.

De uitwerking van de hoofdvraag komt concreet op de volgende deelvragen neer:

1. Hoe goed worden deskundigenrapportages, waarin de conclusie de vorm heeft van een verbale *likelihood ratio*, begrepen door:
  - a. Strafrechters?
  - b. Strafrechtadvocaten?
  - c. Deskundigen van het NFI?
  - d. Hoe en in welke mate verschillen deze drie groepen van elkaar in begrip van de conclusies?
  - e. Welk type denkfouten wordt het meest gemaakt bij interpretatie van de conclusies en verschilt dat tussen de drie onderzoeksgroepen?
2. Hoe goed menen de respondenten de conclusies te begrijpen en in welke mate komt dat vermeend begrip overeen met werkelijk begrip? Is er een kennisoverschatting en hoe groot is deze?
  - a. Verschillen de drie onderzoeksgroepen van elkaar in de verhouding tussen vermeend begrip en feitelijk begrip?
3. In hoeverre kan de mate van begrip van de conclusies worden verklaard door achtergrondkenmerken van de respondenten (zoals leeftijd, ervaring, cursus/specifieke scholing, lezen vakliteratuur)?
4. Leidt een conclusie die op visuele wijze een *likelihood ratio* uitdrukt tot beter begrip dan een conclusie die verbaal een *likelihood ratio* uitdrukt?

## METEN VAN FEITELIJK BEGRIP

Terwijl de meting van vermeend begrip ongewijzigd wordt overgenomen uit Studie 1, is het blok stellingen voor de meting van feitelijk begrip nog verder verbeterd.<sup>25</sup> Dit resulteerde allereerst in het verwijderen van één stelling waarover te veel discussie bleef bestaan bij de begeleidingscommissie. Daarnaast besloten wij, om eerder genoemde redenen, tot het invoegen van een criteriumstelling, bedoeld om met enige zekerheid te kunnen stellen of respondenten als groep niet zomaar wat gegokt hebben bij beantwoording van de stellingen. Deze criteriumstelling (stelling 4, tabel 2), was namelijk pertinent en overduidelijk een incorrecte verwoording van wat de deskundige beoogt te zeggen. In tabel 2 staat het blok van acht stellingen, zoals dat in willekeurige

<sup>24</sup> In de niet-forensische literatuur is veelvuldig onderzoek gedaan naar vergelijking tussen visuele en andere typen schalen bij niet direct meetbare kwantiteiten (zoals geluk of pijn), bijv. Grunberg et al. (1996).

<sup>25</sup> Er is wederom niet gebruikgemaakt van Evetts' hiërarchie van uitspraken, vergelijk voetnoot 15.

volgorde is voorgelegd aan de respondenten na lezing van ieder rapport. De hier weergegeven stellingen hebben betrekking op de conclusie uit het rapport over de beroving op straat (zie tabel 2 en appendix 1). In tabel 2 hebben wij achter de stellingen de correcte antwoorden aangegeven.

Tabel 2 Studie 2: meting van feitelijk begrip (rapportage n.a.v. beroving op straat)

		Is dit een correcte interpretatie van de conclusie?		
		Ja	Nee	WN
1	Er is veel meer dan 50% kans dat de verdachte de persoon op de camerabeelden is.		X	
2	Het is veel waarschijnlijker dat de verdachte de persoon op de camerabeelden is dan dat het iemand anders is die op de camerabeelden staat.		X	
3	De uitkomst van dit onderzoek vormt in sterke mate bewijsmateriaal tegen de verdachte.	X		
4	Er is aangetoond dat de verdachte schuldig is.		X	
5	De resultaten van dit onderzoek zijn erg belastend voor de verdachte.	X		
6	Er zijn treffende overeenkomsten gevonden tussen het gelaat van de verdachte en het gelaat van de persoon op de camerabeelden. Het valt echter niet uit te sluiten dat een aantal andere mensen, net als de verdachte, lijkt op de persoon op de camerabeelden. De verdachte maakt dus deel uit van een groep mensen die allemaal op de beelden zouden kunnen staan. De kans is dus erg klein dat juist de verdachte op de camerabeelden staat.		X	
7	De waarnemingen passen veel beter bij het scenario dat de verdachte dezelfde persoon is als de persoon op de camerabeelden, dan bij het scenario dat de persoon op de camerabeelden iemand anders is.	X		
8	De persoon op de camerabeelden zou ook een andere persoon dan de verdachte kunnen zijn.	X		
Stelling 1 en 2 presenteren verschillende varianten van de prosecutor's fallacy; stelling 4 is de criteriumstelling; stelling 6 verwoordt de defense fallacy.				

Het antwoord ‘weet niet’ rekenen wij fout bij de analyses hierna. Het duidt er immers op dat men in zo’n geval niet in staat is de juiste interpretatie van de conclusie te bepalen.<sup>26</sup> Een laatste globale indicatie van feitelijk begrip

26 ‘Weet niet’ is overigens niet hetzelfde als het geven van het verkeerde antwoord (i.e. een expliciete fout). Onze inzet in deze studie is echter te meten of men conclusies begrijpt. Vanuit dat perspectief rekenen wij ‘weet niet’ gewoon fout, omdat men het kennelijk niet begrijpt. Ter illustratie: vanuit datzelfde perspectief wordt het antwoord ‘weet niet’ dat een student bij een vraag in een schriftelijk tentamen formuleert ook fout gerekend.

verkregen wij aan de hand van een open vraag waarin de respondent werd verzocht de conclusie in eigen woorden weer te geven. Bij de bespreking van de resultaten zullen antwoorden op deze open vraag vooral voor illustratieve doeleinden gebruikt worden.

## MATERIAAL

Het onderzoeksmateriaal voor deze hoofdstudie bestond uit twee fictieve deskundigenrapportages, nagenoeg identiek aan twee van de rapporten in Studie 1. Eén rapport betrof het onderzoek naar de chemische overeenkomsten tussen stukken tape die waren gebruikt om slachtoffers van een *overval op een tankstation* te knevelen en tape die uit de woning van een verdachte was veiliggesteld. Het andere rapport ging over de *beroving op straat* bij een pinautomaat. Het deskundigenonderzoek richtte zich daar op vergelijking van beelden van de bewakingscamera van de pinautomaat met foto's van een verdachte. In appendix 1 staan beide rapporten volledig weergegeven. Ten opzichte van echte rapporten in de praktijk, betreft het hier enigszins verkorte rapportages.

### Box 2 Verbale conclusievorm, beroving op straat

De vraagstelling is geïnterpreteerd als het verzoek om een vergelijkend onderzoek te verrichten en daarbij na te gaan of de bevindingen beter passen bij de hieronder weergegeven hypothese H1 dan bij de (alternatieve) hypothese H2.

Hypothese 1: *De dader van de beroving zichtbaar op SX1\_C1 is dezelfde persoon als verdachte afgebeeld op foto SX1\_B1.*

Hypothese 2: *De dader van de beroving zichtbaar op SX1\_C1 is niet dezelfde persoon als verdachte afgebeeld op foto SX1\_B1.*

(...)

Ik concludeer nu het volgende:

De bevindingen van de hier gerapporteerde gezichtsvergelijking aan de hand van het geselecteerde beeldmateriaal zijn *veel waarschijnlijker* wanneer de daarop afgebeelde persoon dezelfde persoon is (hypothese 1) dan wanneer het een andere persoon betreft (hypothese 2).

Om de verbale vorm van een *likelihood ratio* met een visuele variant te kunnen vergelijken, diende die visuele variant in essentie hetzelfde uit te drukken als zijn verbale tegenhanger. Van ieder deskundigenrapport zijn twee versies

gebruikt: een versie met de thans gebruikelijke verbale *likelihood ratio* en een versie met een visueel uitgedrukte *likelihood ratio*. Ter illustratie staat in box 2 het voor de conclusie relevante deel uit het rapport over de beroving op straat bij een pinautomaat in verbale versie. In box 3 is de visuele variant te zien van de overval op een tankstation.

**Box 3    Visuele conclusievorm, overval tankstation**

De vraagstelling is geïnterpreteerd als het verzoek om een vergelijkend onderzoek te verrichten en na te gaan of de daarbij verkregen resultaten (de bevindingen) beter passen bij het hierna weergegeven Scenario 1 dan bij het alternatieve scenario, Scenario 2.

Scenario 1: *De tape waarmee slachtoffer gekneveld was is afkomstig van de rol tape uit de woning van verdachte.*

Scenario 2: *De tape waarmee slachtoffer gekneveld was is afkomstig van een willekeurige andere rol tape.*

(...)

Ik kom nu tot een conclusie. Deze conclusie geeft aan hoeveel *beter of slechter* mijn onderzoeksresultaten passen bij het scenario dat de tape afkomstig is van de rol tape aangetroffen bij de verdachte (Scenario 1) dan bij het scenario dat de tape afkomstig is van een willekeurige andere rol tape (Scenario 2).

Ik formuleer deze conclusie in de vorm van een verhouding. Deze verhouding is de waarschijnlijkheid dat de verkregen resultaten worden aangetroffen als Scenario 1 het geval is ten opzichte van de waarschijnlijkheid dat de verkregen resultaten worden aangetroffen als Scenario 2 het geval is.

Ik druk deze verhouding uit door een kruisje op onderstaande lijn te plaatsen.

Ze  
er  
sterk  
ten  
gun  
ste  
van  
Scenario 2

Neutraal

Ze  
er  
sterk  
ten  
gun  
ste  
van  
Scenario 1

|

X

De kracht van de conclusie (i.c. de uitgedrukte verhouding van waarschijnlijkheid van de bevindingen onder de twee hypothesen) is in deze studie gevarieerd *tussen* de rapporten. Het onderzoek naar de overval op het tankstation leverde dus zowel in verbale als in visuele conclusievorm een *iets waarschijnlijker* conclusie op, terwijl de beroving op straat in beide varianten van de conclu-

sie in *veel waarschijnlijker* resulteerde. Deze keuze was vooral van praktische aard, namelijk om het onderzoeksdesign niet onnodig complex te maken.

DESIGN

Samenvattend, bestond het materiaal uit twee onderzoeksrapportages; één daarvan (tank) met conclusie *iets waarschijnlijker*, de ander (straat) met conclusie *veel waarschijnlijker*. Van ieder rapport was een conclusievariant met een verbaal uitgedrukte *likelihood ratio* en een variant met een visueel uitgedrukte *likelihood ratio*.

Het onderscheid tussen verbaal en visueel werd niet geschikt geacht om binnen proefpersonen te variëren. Indien een variant beter begrepen wordt dan de andere, zou daarvan namelijk wel eens een leereffect kunnen uitgaan op de beoordeling van die andere variant. Dit werd daarom tussen proefpersonen (*between subjects*) gevarieerd. Binnen proefpersonen (*within subjects*) werd het soort onderzoek waar de deskundige verslag van doet gevarieerd (tank of straat) en daarmee de kracht van de conclusie. Iedere deelnemer aan ons onderzoek kreeg dus twee deskundigenrapporten te beoordelen. Om vertekening van de onderzoeksresultaten door eventuele volgorde-effecten tegen te gaan, is de volgorde waarin beide rapporten aangeboden werden systematisch gevarieerd.

De te benaderen groepen waren strafrechtadvocaten, rechters en raadsheeren werkzaam in de strafsectoren, en deskundigen (van wie sommigen nog in opleiding) van het NFI. De deelnemers aan het onderzoek werden willekeurig toegedeeld aan een van de vier varianten van het materiaal weergegeven in tabel 3. De deelnemers aan het onderzoek werd via een instructieblad voorop de vragenlijst verzocht om na het lezen van het eerste rapport de daarbij gevoegde vragen te beantwoorden alvorens het tweede rapport te lezen.

Tabel 3    Onderzoeksdesign Studie 2

Rechters	►	Tank	(verbaal)	&	Straat	(verbaal)
	►	Straat	(verbaal)	&	Tank	(verbaal)
Advocaten	►	Tank	(visueel)	&	Straat	(visueel)
	►	Straat	(visueel)	&	Tank	(visueel)

## PROCEDURE

Omdat de strategie bij het benaderen van respondenten tussen de drie beroepsgroepen in een aantal opzichten verschilde, worden ze op die punten hierna kort apart behandeld. In iedere beroepsgroep werden potentiële respondenten door ons persoonlijk benaderd, dat wil zeggen met een op naam gestelde brief. Die brief bevatte wat algemene informatie over het onderzoek en natuurlijk het verzoek tot deelname. De deelname was anoniem; er werden geen persoonsgegevens gevraagd en ingevulde vragenlijsten konden in ongemarkeerde portvrije retourenveloppen worden teruggestuurd. Over een periode van twee à drie weken na de mailing van de vragenlijsten zijn twee herinneringsbrieven verstuurd. Vanwege de anonimiteit ontvingen ook degenen die reeds een vragenlijst hadden geretourneerd een herinneringsbrief. In die brief werd dit uitgelegd en werd bedankt voor deelname. Bij de correspondentie met de leden uit alle drie de beroepsgroepen werd een brief van de algemeen directeur van het NFI, dr. T.B.P.M. Tjin-A-Tsoi, gevoegd. In deze brief werd gewezen op het belang van het onderzoek en werd de geadresseerde verzocht medewerking te verlenen.

## WERVING STRAFRECHTADVOCATEN

De meeste strafrechtadvocaten in Nederland zijn lid van de Nederlandse Vereniging van Strafrecht Advocaten (NVSA). Op de website van deze vereniging is de ledenlijst te raadplegen.<sup>27</sup> In september 2008 zijn alle leden van de NVSA benaderd.

## WERVING RECHTERS

Na een gemotiveerd verzoek verleende de Raad voor de rechtspraak toestemming om onderzoek te verrichten bij rechters en raadsheren in de strafsectoren. In overleg met de Raad zijn twee gerechtshoven en vijf arrondissementsrechtbanken geselecteerd.<sup>28</sup>

In een brief aan de presidenten van de geselecteerde rechtbanken en hoven werd het onderzoek aangekondigd en kort toegelicht (bijgesloten was ook de toestemmingsbrief van de Raad voor de rechtspraak en de brief van de directeur NFI). Kort daarop werden de voorzitters van de strafsectoren bij die

27 Wij danken het bestuur van de NVSA voor zijn hulp bij het beschikbaar stellen van de ledenlijst in een voor een adresseringsprogramma bruikbare vorm.

28 Deze selectie was ingegeven door twee overwegingen: enerzijds regionale spreiding en anderzijds praktische overwegingen van de zijde van de Raad in verband met ander, niet gerelateerd onderzoek bij rechters dat tegelijkertijd liep.



gerechten benaderd met het verzoek de leden van hun sector te attenderen op het op handen zijnde onderzoek en hen te verzoeken medewerking te verlenen. Pas daarna, in oktober 2008, zijn de vragenlijsten naar de rechters en raadsheren bij de zeven gerechten verstuurd.

## WERVING DESKUNDIGEN

In een interne e-mail aan alle deskundigen en deskundigen in opleiding bij het NFI, kondigde mw. dr. M. Sjerps (statisticus bij het NFI) het onderzoek aan. Tevens gaf zij de NFI-medewerkers daarbij de gelegenheid hun instemming te onthouden aan het gebruik van het privéadres voor de mailing van de vragenlijsten. Het gebruik van de privéadressen vonden wij voor deze groep uitermate belangrijk, omdat het onderwerp een essentieel onderdeel van hun dagelijks werk betreft én omdat de meeste deskundigen op het NFI kantoorruimte met elkaar delen. Slechts drie medewerkers maakten bezwaar tegen toezending op het huisadres, zij kregen de vragenlijst toegestuurd op het werk. Op het moment van versturen van de vragenlijsten werd door mw. Sjerps in een bericht aan alle afdelingshoofden verzocht de medewerkers te enthousiasmeren voor dit onderzoek. De vragenlijsten zijn in september 2008 verstuurd aan alle deskundigen en deskundigen in opleiding van het NFI.<sup>29</sup> Hierna wordt, tenzij anders vermeld, met 'deskundigen' bedoeld de deskundigen die tekenbevoegdheid hebben bij de rapportages van het NFI, alsmede de deskundigen in opleiding.<sup>30</sup>

## RESPONS

In tabel 4 staan de kerngegevens van dataverzameling en respons bij de drie door ons benaderde groepen professionele procesdeelnemers.

*Tabel 4 Respons Studie 2: strafrechtadvocatuur, zittende magistratuur, NFI*

	Vragenlijst verzonden	Benaderd	Ontvangen	Respons
Strafrechtadvocaten	1 september 2008	361	69	19%
Rechters en raadsheren	7 oktober 2008	260	118	45%
Deskundigen (i.o.)	24 september 2008	145	99	68%

29 Onmiddellijk nadat de tweede herinneringsbrief was verstuurd, zijn de door ons gebruikte namen en adressen van deskundigen vernietigd.

30 De groep deskundigen in opleiding betreft 28% van onze respondenten, zoals hierna zal blijken.

Terwijl in absolute zin de aantallen per beroepsgroep voldoende zijn voor de beoogde statistische analyses, laat tabel 4 tussen de beroepsgroepen verhoudingsgewijs grote verschillen in respons zien. Met een vijfde respons hebben we het minste medewerking gekregen van advocaten. Bovendien is 19% sowieso een lage respons te noemen. Dit is een belangrijke kanttekening bij het generaliseren van de onderzoeksbevindingen over de advocaten. Het responspercentage bij de rechters en raadsheren van de zeven geselecteerde gerechten is zonder meer bevredigend. Met enige voorzichtigheid kan dit geïnterpreteerd worden als een teken dat het een onderwerp betreft dat door de zittende magistratuur als belangrijk en urgent wordt beschouwd.

Ruim twee derde van de benaderde NFI-medewerkers deed mee aan het onderzoek. Dat is in algemene zin een hoge respons te noemen. De pessimist zal er echter op wijzen dat het opvallend is dat een derde van de deskundigen niet meedoet, waar het onderzoek toch voor velen onder hen de kern van hun eigen werk raakt.



## Resultaten Studie 2<sup>31</sup>

### CRITERIUMSTELLING

Alvorens dieper in te gaan op door ons gemeten feitelijk begrip, behandelen we kort de antwoorden op de criteriumstelling. Als velen zouden instemmen met de stelling *Er is aangetoond dat de verdachte schuldig is*, kan ervan uitgegaan worden dat men niet zorgvuldig te werk ging bij het beantwoorden van de vragenlijst.

Van alle deelnemers aan het onderzoek gaven er drie (1%) aan dat deze stelling correct is in het geval van de tapesvergelijking (tankstation) en twaalf (4%) bij de vergelijking van het beeldmateriaal (straat). Dat aantal is zo gering dat wij concluderen dat de respondenten over het algemeen de vragenlijst serieus invulden. De respondenten zijn allen in de analyse meegenomen.

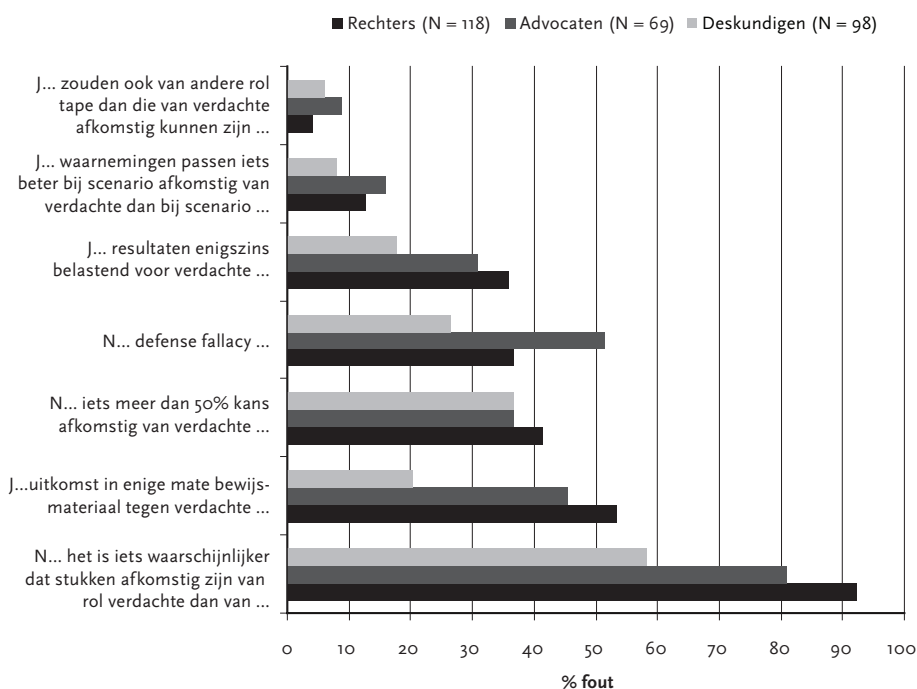
### FEITELIJK BEGRIP: INDIVIDUELE STELLINGEN

In figuur 4 en in figuur 5 is voor, respectievelijk, de beroving van het tankstation en de beroving op straat het percentage foute antwoorden per stelling weergegeven. De weergave is verder gedifferentieerd naar beroepsgroep. Uit deze figuren is op te maken hoe moeilijk de verschillende stellingen waren om correct te classificeren en in welke mate de drie beroepsgroepen van elkaar verschilden in aantallen fouten. Voor de duidelijkheid staat in de figuren bij iedere stelling het correcte antwoord aangeduid (J of N). Voor de volledige tekst van de stellingen, zie tabel 2.

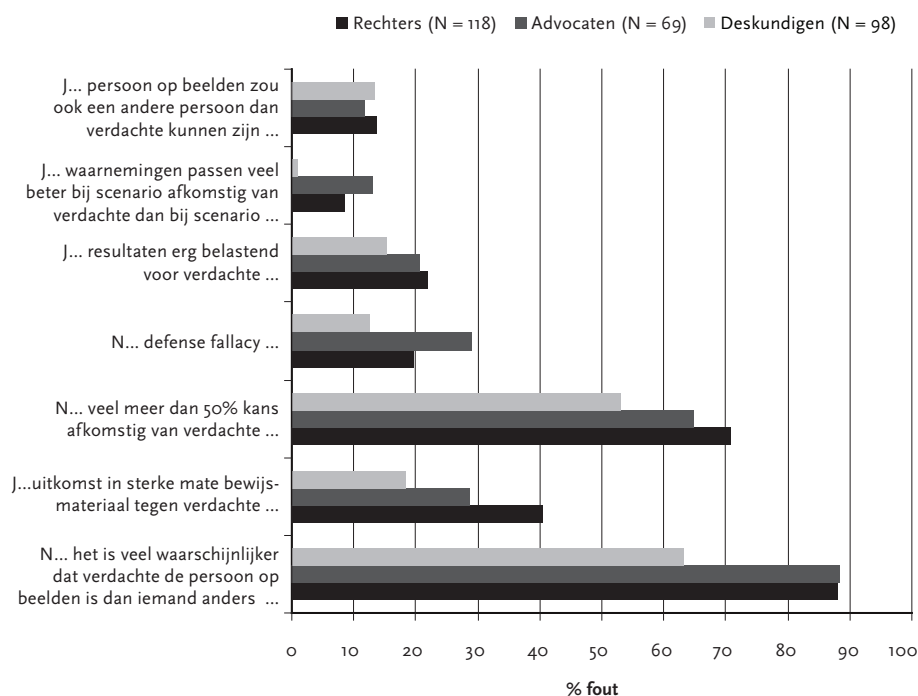
Zowel figuur 4 als figuur 5 laat grofweg dezelfde patronen en verschillen zien. Het wordt onmiddellijk duidelijk dat sommige stellingen tot veel meer fouten hebben geleid dan andere. Evenals in Studie 1 levert de beruchte *prosecutor's fallacy* bij iedere beroepsgroep de meeste fouten op. Van de rechters en de advocaten trapt (soms veel) meer dan 80% in beide casus in deze val (onderste stelling in de figuren). Maar ook de deskundigen zelf gaan hier op grote schaal de mist in. Bij het onderzoek naar de tapes (tankstation) meent 58% van de deskundigen ten onrechte dat het iets waarschijnlijker is dat de stukken tape afkomstig zijn van de rol van verdachte dan van een andere rol (figuur 4). Bij het

<sup>31</sup> In appendix 2 zijn de relevante cijfers en statistische onderbouwing terug te vinden. In de hoofdtekst worden, evenals bij de bespreking van Studie 1 hiervoor, de resultaten vooral aan de hand van figuren gepresenteerd.

beeldvergelijkend onderzoek in verband met de beroving op straat maakt 63% van de deskundigen deze fout. Maar ook in de wat andere formulering van de prosecutor's fallacy, *Er is veel meer dan 50% kans dat de verdachte de persoon op de camerabeelden is* (figuur 5) gaat de meerderheid van de deskundigen in de fout. Een andere stelling waar veel deskundigen kennelijk moeite mee hebben, is de *defense fallacy*, die bij de vergelijking van tapes (tankstation) meer dan een kwart van de deskundigen tot het verkeerde antwoord verleidt (zie figuur 4). Naast de constatering dat sommige stellingen door verrassend veel deskundigen verkeerd geïnterpreteerd worden, laten de figuren ook zien dat in beide casus op bijna alle stellingen van de drie beroepsgroepen de deskundigen de minste foute antwoorden geven. Dat is in ieder geval geruststellend, maar bij de betekenis van de absolute hoeveelheid misclassificaties onder de groep deskundigen zullen we hierna nog uitvoerig stilstaan. Een ander opvallend aspect in beide figuren, maar vooral in figuur 4, is de verdeling van *defense fallacies* tussen de drie groepen. Dat de *defense fallacy* vooral zou passen in de gedachtegang van advocaten wordt hier zonder meer bevestigd. Deze denkfout doet zijn naam dan ook eer aan. Met 52% *defense fallacies* in figuur 4 laten de advocaten de twee andere groepen ver achter zich.



Figuur 4 Feitelijk begrip per stelling, Studie 2, beroving tankstation: rechters, advocaten en deskundigen



*Figuur 5 Feitelijk begrip per stelling, Studie 2, beroving op straat: rechters, advocaten en deskundigen*

Ten slotte, door iedere groep worden relatief weinig fouten gemaakt betreffende de constatering dat de stukken tape ook van een andere rol kunnen zijn en dat de persoon op de beelden ook een andere persoon kan zijn. Kennelijk is voor de overgrote meerderheid binnen iedere beroepsgroep na het lezen van de rapportage en conclusie die mogelijkheid niet geheel afgesloten. Ook de stelling die de conclusie van de deskundige in de rapporten nog het meest direct parafraseert (waarnemingen passen iets beter bij het ene dan bij het andere scenario), wordt ook als zodanig herkend en door verreweg de meeste respondenten terecht als correct geclassificeerd. Dat is natuurlijk hoopvol te noemen, omdat deze laatste stelling precies de *raison d'être* verwoordt van het rapporteren in termen van aannemelijkheidsverhoudingen. Daartegenover staat echter dat sommige andere stellingen, die ook een correcte verwoording vormen, tot soms meer dan de helft foute antwoorden leiden. Wij interpreteren dat als een signaal dat veel respondenten toch niet boven de stof staan.

## OPEN VRAGEN

Voorafgaand aan de stellingen bij ieder rapport werd de respondenten telkens gevraagd de conclusie van de deskundige in hun *eigen woorden* te formuleren. Bij alle drie groepen, rechters, advocaten en deskundigen, zien we zeer regelmatig dat ook in deze eigen bewoordingen de *prosecutor's fallacy* wordt verwoord. In box 4 staan enkele voorbeelden, uit vele, van alle groepen dooreen. Het is zeker niet zo dat slechts door het aanbieden van 'valkuilstellingen' de respondenten tot het maken van fouten worden geprovoceerd. Uiteraard zijn er ook voorbeelden te geven van een smetteloze verwoording van de conclusies, maar het aantal fouten als zojuist behandeld, domineert.

### Box 4 Conclusie in eigen woorden: voorbeelden van *prosecutor's fallacy*

- *De beide stukken tape zijn zeer waarschijnlijk afkomstig van de rol tape uit de woning van verdachte.*
- *De tape gebruikt bij de overval is >50% dezelfde als de tape gevonden in de woning.*
- *De tape waarmee het slachtoffer is gekneveld, is waarschijnlijk afkomstig van de rol tape die van de woning van de verdachte is aangetroffen.*
- *Een zekere mate van waarschijnlijkheid wijst uit dat de tape waarmee delict werd gepleegd afkomstig is van tape aangetroffen in woning van verdachte.*
- *Er is IETS meer kans dat de verdachte de dader is dan dat de verdachte de dader niet is.*
- *Het is hem.*
- *Het is waarschijnlijker dat de gebruikte tape komt van de rol van verdachte dan van een andere rol is groter dat tape wel van verdachte was dan dat dat niet het geval was.*
- *Op basis van de resultaten van het vergelijkend onderzoek wordt geconcludeerd dat scenario 1 iets waarschijnlijker is dan scenario 2.*
- *Op basis van vergelijkend onderzoek is hypothese 1 iets waarschijnlijker dan hypothese 2.*

Wat door de bestudering van de antwoorden op de stellingen niet aan de orde komt, is dat respondenten zich plegen te ergeren aan de rapportages als geheel, zonder dat dat zich onmiddellijk in verwarring over de conclusie in termen van de aannemelijkheidsverhouding uit. Dat blijkt uit hetgeen in de vrije verwoording van de conclusies naar voren wordt gebracht. Er klinkt allerlei kritiek door op de onderzoeksmethode of anderszins: de lezers willen het onderzoek anders doen of achten de conclusie te omslachtig en onnodig voorzichtig. Enkele treffende verwoordingen hiervan staan weergegeven in box 5. Ten slotte blijkt uit de vrije verwoording ook nogal eens dat de respondent eigenlijk niet weet wat hij met de rapportage aan moet, maar ook (het andere uiterste) dat de respondent zich niet kan voorstellen dat je iets anders zou

kunnen zeggen dan het rapport doet. Sommige respondenten hanteren vage bewoordingen bij de eigen formulering (box 6). Een en ander illustreert dat ook buiten het onderwerp van dit onderzoek, de formulering van waarschijnlijkheidsoordelen, deskundigenrapportages onderwerp zijn van onduidelijkheid en kritiek.

### Box 5 Conclusie in eigen woorden: voorbeelden van kritiek op het rapport

- *Dit onderzoek was zinloos. Ga zelf naar wat winkels en bouwmarkten om te tellen hoeveel % van de rollen tape BLAUW is.*
- *Eerst maar eens onderzoek naar de aangetroffen verontreinigingen. Lijkt me sterk dat er niet een schoon stukje tape te vinden is! Het is een lijmlaag die niet in de database zit, maar onbekend is hoe groot die database is. Eerst maar langs de bouwmarkten op zoek naar blauwe tape om de zeldzaamheidswaarde van deze tape te bepalen. Vooralsnog kun je er niets mee.*
- *Hoewel er geen zekerheid is in wetenschappelijke zin, is het in gegeven voldoende zekerheid om als bewijs te dienen. Beter dan een 'gewone' herkenning, maar als lezer van zo'n rapport denk je wel: 'Doe in godsnaam niet zo moeilijk!' Je kan te ver gaan in het vertalen van ervaringsfeiten in wetenschappelijk bewijs. Hoe loffelijk het streven ook is, je moet niet de suggestie willen wekken dat recht wetenschap is.*
- *Ik begrijp dat er een aantal gelaatskenmerken bij de dader en bij de verdachte hetzelfde zijn en dat deze combinatie van kenmerken vrij zeldzaam is. Waarom dan niet de conclusie dat de kans dat ...*
- *Ik heb alleen maar vragen bij dit rapport. Bij rapporten als deze plegen allerlei berekeningen en herkenningspunten te worden gevoegd. Hier heb ik alleen 'een veelheid aan overeenkomsten in morfologisch en antropologische kenmerken'. Geen idee waarom deze bevindingen veel waarschijnlijker zijn wanneer de dader dezelfde persoon is als de verdachte dan wanneer het een andere persoon is.*
- *Of ingetogener in eerste stuk bij Resultaten, of zwaarder in Conclusie.*

### Box 6 Conclusie in eigen woorden: moeite met herformuleren

- *Conclusie spreekt voor zich.*
- *Dat kan ik niet in eigen woorden weergeven.*
- *Er is een goede kans dat de gebruikte tape afkomstig is van de rol die bij verdachte is aangetroffen, maar misschien ook niet ...!*
- *Het zou zo kunnen zijn dat de tape waarmee het slachtoffer gekneveld was afkomstig is van de rol uit de woning van de verdachte.*
- *'Hij lijkt er veel op.'*
- *Ik zie zo gauw geen alternatief.*
- *Nee, onvoldoende info.*
- *Nietszeggend.*



## GESOMMEERDE SCORES ALS METING VAN FEITELIJK BEGRIIP

Uit Studie 1 bleek dat er weinig samenhang was tussen de antwoorden op de uiteenlopende stellingen. In deze hoofdstudie met rechters, advocaten en deskundigen lijkt daar aan de hand van het iets gewijzigde blok stellingen enige verbetering in te zijn gekomen. Wanneer we de twee casus (tank en straat) samennemen<sup>32</sup> en vervolgens voor de drie te onderscheiden beroepsgroepen apart een sommering van het aantal goede antwoorden maken, blijkt dat voor de rechters en voor de deskundigen een naar algemene criteria intern voldoende consistente schaal op te leveren. Bij de advocaten laat die interne consistentie te wensen over en valt deze beneden de gebruikelijke normen daarvoor.<sup>33</sup> Het berekenen van interne consistentie van een schaal voor de twee casus tezamen en de drie beroepsgroepen tezamen levert wederom een acceptabele schaal op (Cronbachs  $\alpha = 0,67$ ).

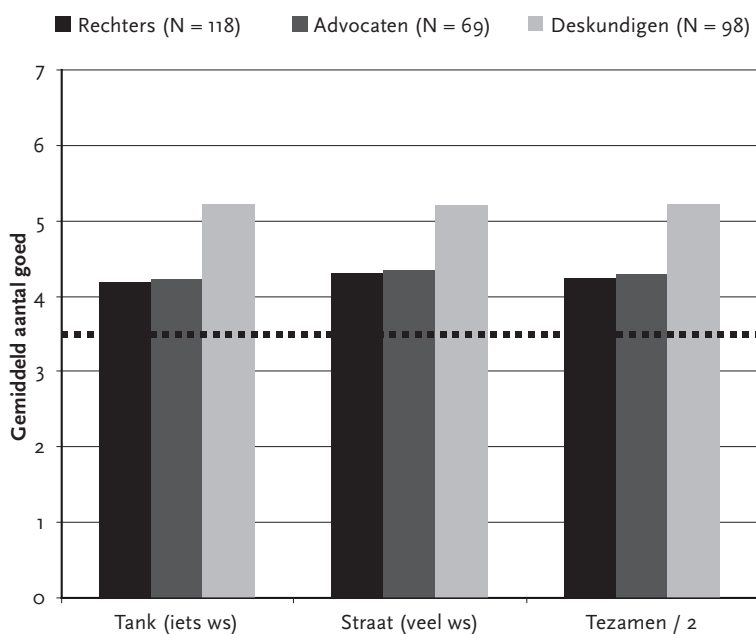
Voor twee van de drie groepen en voor alle groepen tezamen vormen de antwoorden op de stellingen kennelijk een schaal met voldoende interne consistentie naar statistische maatstaven. De advocaten vormen een uitzondering. Het behoort tot de mogelijkheden om apart voor de advocaten een intern wat consistentere schaal te construeren door sommige stellingen daarbuiten te laten. Dat zou echter de vergelijkbaarheid van de antwoorden van advocaten en anderen geweld aandoen. Daar wij met het oog op schaalconstructie primair geïnteresseerd zijn in het gehele blok stellingen tezamen, hebben wij besloten om voor alle respondenten, dus ook voor advocaten, één en dezelfde schaal te hanteren. Die schaal is dus simpelweg de optelsom van het aantal goede antwoorden. Anders gesteld, de schaal komt overeen met een rapportcijfer dat bestaat uit het totaal aantal punten waarbij voor iedere vraag een goed antwoord een punt oplevert. Bij de sommering van goede antwoorden per deskundigenrapport betekent dit een slechtst haalbare score van 0 en een maximale score van 7, als alle stellingen (met uitzondering van de criteriumstelling) juist werden geklasseerd.

32 Hierna zal blijken dat het samenvoegen van die antwoorden zonder meer verantwoord is.

33 Cronbachs  $\alpha$  somschaal aantal goed (14 items): rechters, 0,62; advocaten, 0,39; deskundigen, 0,72. Wij merken op dat het wellicht mogelijk is onderliggende dimensies te identificeren die schuilgaan achter deze aanpak met één somschaal. Die exercitie laten wij hier achterwege, omdat het doel nadrukkelijk was om een en dezelfde naar statistische maatstaven acceptabele schaal te construeren, waarop de verschillende groepen met elkaar kunnen worden vergeleken.

## FEITELIJK BEGRIP, PER CASUS EN TEZAMEN

De drie beroepsgroepen worden nu niet langer op individuele stellingen met elkaar vergeleken, maar op totaalscores per casus en voor de twee casus gezamenlijk (gemiddeld). Figuur 6 geeft dit gemiddeld aantal goede antwoorden van de drie beroepsgroepen.<sup>34</sup> In de figuur geeft de stippellijn het niveau aan dat men zou verwachten als respondenten voor het antwoord op iedere stelling een munt zouden opgooien. Dat niveau ligt dus op het schaal midden van  $3\frac{1}{2}$ .



*Figuur 6 Feitelijk begrip, Studie 2: gemiddeld aantal correct geclassificeerde stellingen*

Figuur 6 laat zien dat iedere beroepsgroep beter presteert dan alleen op basis van het opgooien van een munt te verwachten is. Voor de rechters en de advocaten is daarmee dan wel het meeste gezegd. Weliswaar spreken we van statistisch significante verschillen ten opzichte van het schaal midden,<sup>35</sup> maar erg ver daarboven komen de gemiddelden niet. Over de twee casus heen komen rechters gemiddeld op 4,25 goed geclassificeerde stellingen en advocaten

<sup>34</sup> Zie tabel A2.7 in appendix 2, voor de precieze cijfers, standaarddeviaties en statistische toetsen.

<sup>35</sup> Zie noot bij tabel A2.7 in appendix 2.

op 4,28. Met de prestaties op de individuele stellingen in het achterhoofd, mag dit weinig verbazing wekken.

Deskundigen presteren gemiddeld beter met scores van net boven de 5,2. Dat deskundigen het aanmerkelijk beter doen dan de andere beroepsgroepen blijkt ook uit vergelijking van de percentages deelnemers die een 6 of het maximum van 7 scoren, zoals weergegeven in tabel 5. Men kan erover twisten hoeveel vragen een persoon goed moet hebben, willen we hem klasseren als iemand die het rapport begrijpt. Onzes inziens kunnen we iemand die slechter scoort dan het verwachte aantal goede antwoorden dat het opgooien van een munt zou produceren (i.c. 3,5), classificeren als behept met onbegrip. Maar het gaat ons wat ver om iedereen die het beter doet dan een munt als begripsvol te karakteriseren. Gezien het belang van deskundigenrapportages voor de rechtspraak vinden wij de lat dan wel erg laag liggen. Wij zullen, arbitrair, daarom wie '4' of '5' van de zeven antwoorden goed heeft, kenschetsen als 'enig begrip' hebbend, en slechts wie bijna alles (6) of alles (7) goed heeft, toont naar onze maatstaven 'goed begrip'.

*Tabel 5 Feitelijk begrip, Studie 2: percentages per categorie aantal goed (kolompercentages)*

Aantal goed	Rechters (N = 118)		Advocaten (N = 69)		Deskundigen (N = 98)	
	tank	straat	tank	straat	tank	straat
0-3 onbegrip	31	22	20	16	14	12
4-5 enig begrip	51	70	70	78	38	46
6-7 goed begrip	18	9	10	6	48	42

Bij de deskundigen ligt het percentage dat 'goed begrip' toont in de tankcasus op 48% en in de straatcasus op 42%. Bij de rechters en advocaten is dat respectievelijk 18% en 10% (tank) en 9% en 6% (straat). Wederom zou men zich kunnen afvragen of de score van deskundigen ook in absolute zin goed genoeg te noemen is; daar komen we nog op terug. In plaats van naar de beste scores, kan in tabel 5 natuurlijk ook naar de slechtste scores worden gekeken. Zo heeft 31% van de rechters niet meer dan drie stellingen goed beantwoord in de tankcasus. Datzelfde geldt voor 20% van de advocaten en zelfs voor 14% van de deskundigen.

## GEEN WEZENLIJK VERSCHIL VAN FEITELIJK BEGRIIP TUSSEN DE TWEE CASUS

Tussen de feitelijke begripscores op basis van de stellingen in de casus naar aanleiding van de beroving van het tankstation en de casus naar aanleiding van de beroving op straat is geen significant verschil te constateren. Voor ieder van de drie beroepsgroepen stelden wij vast dat de conclusies in beide rapporten even goed (of slecht) begrepen worden (zie tabel A2.8 in appendix 2). Om die reden wordt hierna het gemiddelde van de begripscores over deze twee rapporten tezamen aangegrepen voor verdere analyses. Waar inhoudelijk relevant wordt natuurlijk naar de afzonderlijke rapporten verwezen.

## INTERMEZZO: OVER FEITELIJK BEGRIIP

Het beeld dat zich nu opdringt over de mate van feitelijk begrip van verbale aannemelijkheidsverhoudingen in deskundigenrapportages is niet rooskleurig. Door de consumenten van dergelijke rapporten in de rechtszaal, die in ons onderzoek participeerden, worden de conclusies van de forensisch-technisch deskundige in de twee fictieve casus niet veel beter begrepen dan men op basis van kans kan verwachten. De meest beruchte valkuilen bij interpretatie van dergelijke conclusies blijken daarbij welhaast onweerstaanbaar. Deze stand van zaken vinden wij zorgelijk, omdat een van de bouwstenen voor de bewijsbeslissing van rechters (in voorkomende zaken) kennelijk niet goed wordt begrepen. Deze constatering staat geheel los van de vraag of er dientengevolge dan ook meer foute beslissingen worden genomen. Daar hebben wij geen onderzoek naar gedaan.

Nu betreft het hier een wetenschappelijk quasi-experimenteel onderzoek met fictieve casus en in die context hebben bepaalde interpretatiefouten geen repercussies voor direct betrokkenen in zo'n zaak: er zijn hier immers geen betrokkenen van vlees en bloed. In de praktijk is dat natuurlijk anders. Weliswaar kan de rechter zich dan laten voorlichten of laten assisteren door collega's en kan hij zo nodig een deskundige op zitting oproepen om een rapport en de daarin verwoorde conclusie toe te lichten. Maar de vraag blijft of de rechter deze toelichting dan wel begrijpt en of zijn collega's niet in hetzelfde schuitje zitten. Bovendien, en dat durven we nu voorzichtig hardop te zeggen, is het dan nog de vraag of de opgeroepen deskundige in kwestie de materie wel zo goed beheerst om een en ander adequaat uit te leggen.

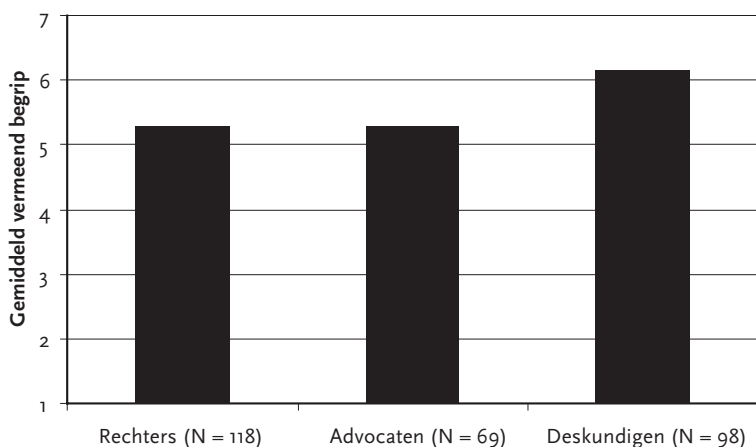
Er tekent zich hier, kortom, een duidelijk probleem af. Het komt ons weinig zinvol voor om misverstanden bij de interpretatie van verbale *likelihood ratio's* te wijten aan eventuele tekortkomingen van de specifieke groepen professionele lezers van deze rapporten. Het probleem lijkt eerder besloten te liggen in het concept *likelihood ratio* zelf en de uitleg daarvan. En dat is een probleem waar ook de deskundigen van het NFI mee worstelen. Op dit punt is

het goed te memoreren aan de resultaten van Studie 1, waar bleek dat mathematisch hoog geschoolde mensen de plank even vaak missloegen als juristen. Algemene (bij)scholing in wiskunde en statistiek lijkt dan ook niet de meest voor de hand liggende oplossing voor de hier geconstateerde situatie.

Hierna presenteren we de gedachten van de drie professionele groepen zelf, over de vraag tot wier domein problemen bij interpretatie van conclusies van deskundigen zouden behoren. Het probleem kan eerst nog verder gekwalificeerd worden. In welke mate menen de deelnemers aan het onderzoek dat ze de conclusies hebben begrepen? Is er bij diezelfde personen die interpretatiefouten maken een zelfbewustzijn van de situatie of is er sprake van een kennisoverschatting?

### VERMEEND BEGRIP EN FEITELIJK BEGRIP: EEN KENNISOVERSCHATTING?

Na het lezen van ieder rapport beantwoordde de respondent de vraag hoe goed hij of zij de conclusie begreep aan de hand van dezelfde 7-puntsschaal als gebruikt in Studie 1. Een maximale score van 7 duidt op volledig begrip. In figuur 7 staan de (gemiddelde) zelftoegekende vermeend-begripsscores van de drie beroepsgroepen weergegeven.<sup>36</sup>



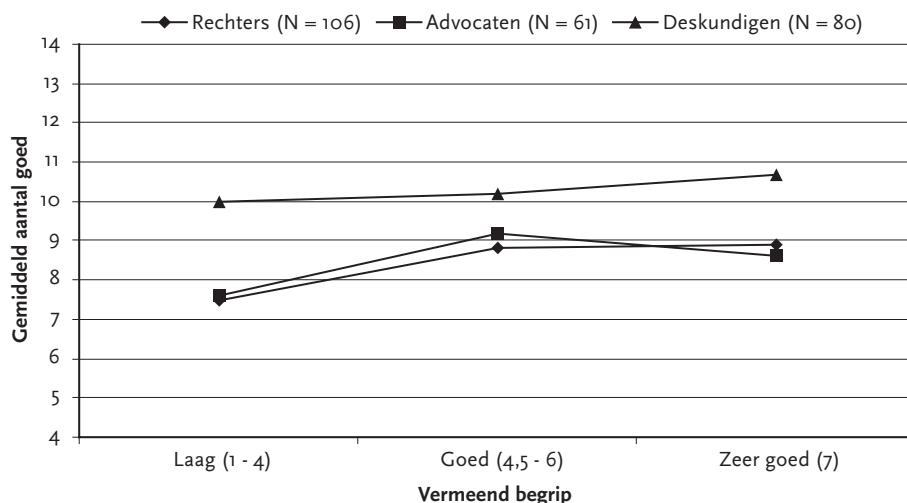
*Figuur 7 Vermeend begrip, Studie 2: gemiddelde score op schaal van vermeend begrip (tank en straat tezamen)*

Figuur 7 laat een soortgelijk beeld zien als vermeend begrip in Studie 1. De beoordelaars van de rapporten menen zelf in het algemeen de conclusies goed tot zeer goed te begrijpen. Rechters en advocaten hebben eenzelfde niveau van

<sup>36</sup> Wij maken, zoals reeds opgemerkt, nu geen onderscheid tussen de twee casus, omdat daar noch in feitelijk, noch in vermeend begrip sprake was van verschil.

vermeend begrip; gemiddeld 5,3. De deskundigen bevinden zich als groep zelfs nog dichterbij het schaalmaximum (gemiddeld 6,2). Niet minder dan 85% van alle deskundigen heeft zichzelf een 6 of 7 op deze schaal gegeven. Men meent de conclusies in de rapporten goed te hebben begrepen, terwijl, zoals in de vorige paragraaf bleek, het feitelijk begrip die mening niet zonder meer ondersteunt.

Het gebrek aan samenhang tussen vermeend begrip en feitelijk begrip wordt in figuur 8 per beroepsgroep verder geïllustreerd. De gemiddelde scores voor vermeend begrip zijn daartoe verdeeld in drie categorieën: laag vermeend begrip (minder dan 4 op de schaal van vermeend begrip), goed vermeend begrip (4-6) en zeer goed (7). Feitelijk begrip is in deze figuur de optelsom van alle goed geclassificeerde stellingen over de twee rapporten heen (maximumscore is dus 14). Figuur 8 laat binnen de beroepsgroepen slechts matige samenhang zien tussen vermeend begrip en feitelijk begrip. Met enige moeite kan men uit figuur 8 opmaken dat de rechters en advocaten die van zichzelf menen de conclusies slecht te hebben begrepen, dat ook met realiteitszin doen.<sup>37</sup> Dat bij de deskundigen geen samenhang waarneembaar is, hoeft niet te verbazen. De meeste van hen (85%) schurken immers al tegen het schaalmaximum van vermeend begrip aan. Er is dan nauwelijks ruimte voor enige samenhang tussen variatie in vermeend begrip en variatie in feitelijk begrip.<sup>38</sup>



*Figuur 8 Vermeend begrip met feitelijk begrip, Studie 2: gemiddelden per beroepsgroep*

37 De verschillen zijn echter klein en slechts significant op toetsniveau van 5%. Zie tabel A2.10. Vanwege de meervoudige toetsen (en dus risico op kanskapitalisatie) menen wij dat dit onvoldoende is om er verder waarde aan te hechten.

38 De (Spearman's rho) correlatiecoëfficiënt voor de samenhang tussen beide schalen is dan ook nagenoeg 0 (tabel A2.10).

Kennelijk is er sprake van een flinke kennisoverschatting. Die kennisoverschatting is het omvangrijkst bij de rechters en advocaten. Het niveau van feitelijk begrip bij deze twee groepen rechtvaardigt geenszins hun zelfbeleden begrip van de conclusies in de twee deskundigenrapporten. Voor de deskundigen geldt dat het verschil weliswaar kleiner is, maar ook voor deze groep kunnen we constateren dat de mate van feitelijk begrip niet in verhouding staat tot het blakend zelfvertrouwen waarvan de zeer hoge scores op vermeend begrip een expressie zijn. Bovendien is het in dezen zonder meer gerechtvaardigd om aan het begrip van de conclusie als aannemelijkheidsverhouding bij deskundigen hogere eisen te stellen.

### VERTROUWDHEID MET KWANTITATIEF MATERIAAL IN HET ALGEMEEN

Aan alle respondenten is gevraagd naar algemeen zelfvertrouwen bij het omgaan met cijfers en statistisch materiaal. De vraag luidde: *Hoe vertrouwd voelt u zichzelf als het gaat om het lezen en begrijpen van cijfermateriaal en statistische gegevens?*, waarop het antwoord kon worden gegeven op een 7-puntsschaal van 1 ('ik voel me helemaal niet vertrouwd in het omgaan met cijfers') tot 7 ('ik voel me heel erg vertrouwd in het omgaan met cijfers'). De samenhang is onderzocht van dit algemeen kwantitatief zelfvertrouwen met enerzijds feitelijk begrip en anderzijds vermeend begrip. Bij rechters en advocaten is geen sprake van samenhang tussen deze maten. Bij de deskundigen hangt het goede kwantitatief zelfvertrouwen in geringe mate samen met feitelijk begrip en met vermeend begrip. Met beide vormen van begrip is er een lichte samenhang (in beide gevallen: Spearman's  $\rho = 0,23$ ,  $p < 0,05$ ).

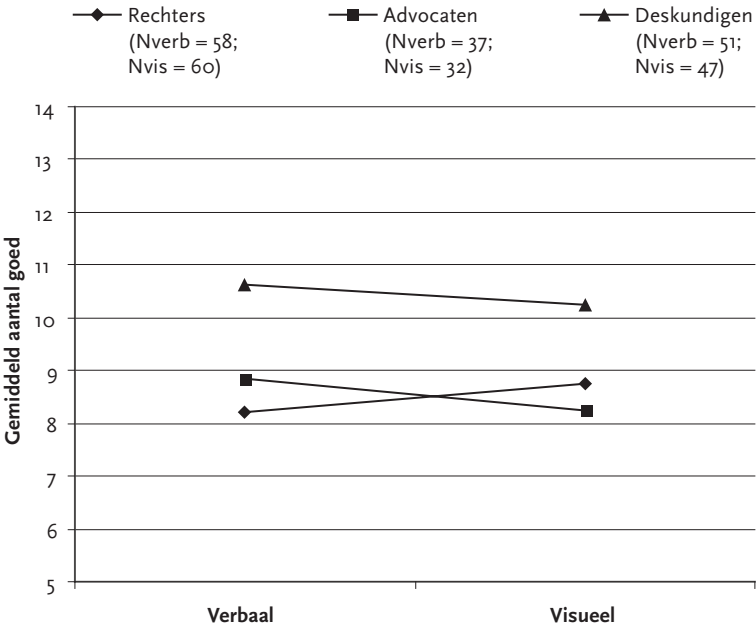
### WORDT EEN VISUEEL UITGEDRUKTE LIKELIHOOD RATIO BETER BEGREPEN?

De vierde onderzoeksvraag luidde of een conclusie die op visuele wijze een *likelihood ratio* uitdrukt tot beter begrip leidt dan een conclusie die verbaal een *likelihood ratio* uitdrukt. In box 3 (hiervoor) staat een voorbeeld van een visuele variant. Van beide rapporten in Studie 2 was er een variant met een verbaal uitgedrukte *likelihood ratio* in de conclusie en één met een visueel uitgedrukte *likelihood ratio* (zie ook tabel 3).

In figuur 9 is voor ieder van de drie beroepsgroepen de mate van feitelijk begrip voor de verbale variant en de visuele variant weergegeven. Figuur 9 laat bij iedere beroepsgroep zien dat het voor feitelijk begrip niet uitmaakt of de conclusie in de gebruikelijke verbale aannemelijkheidsverhouding gegoten is, of in de door ons ontwikkelde visuele vorm.<sup>39</sup> Overigens geldt dat onver-

39 Op individuele stellingen in de afzonderlijke casus werden ook geen significante verschillen aangetroffen tussen verbaal en visueel.

minderd ook voor het vermeend begrip (niet afgebeeld, doch zie tabel A2.11 in appendix 2). Daarmee, en met de vaststelling dat het er niet slechter op wordt, is dan ook alles gezegd over de visuele variant.



Figuur 9 Feitelijk begrip voor verbale en visuele conclusievariant, Studie 2

**KWANTITATIEVE VOORSTELLING BIJ VERBALE LABELS AANNEMELIJKHEIDS-VERHOUDING**

Aan de proefpersonen die de verbale variant van de conclusies voorgeschoteld kregen, is een additionele vraag gesteld over de kwantitatieve vertaling van de labels van de schaal, die liep van ‘ongeveer even waarschijnlijk als’ tot ‘zeer veel waarschijnlijker’.<sup>40</sup> Bijna de helft van de respondenten heeft desgevraagd geen behoefte aan een kwantitatieve voorstelling bij de verbale labels (voor alle beroepsgroepen). Degenen die zich wel een kwantitatieve parallel voorstellen, is gevraagd te kiezen uit een reeks mogelijkheden. Deskundigen kiezen relatief vaak voor de reeks 1-10-100-1000-10000 (33%), rechters en advocaten relatief vaak voor 1-2-3-4-5 (ruim 20%). Een bescheiden percentage, 14% heeft een eigen vertaling gegeven, maar daar konden wij weinig lijn in ontdekken.

40 Zie appendix 3 voor de exacte vraagstelling.



## BEGRIP VERKLAREN

Het ontwerp van deze studie was vooral gericht op het meten en in kaart brengen van feitelijk begrip en van een eventuele kennisoverschatting bij de professionele procesdeelnemers. Desalniettemin betreft de derde onderzoeksvraag de mate waarin verschillen in begrip van aannemelijkheidsverhoudingen in conclusies verklaard kunnen worden aan de hand van enkele achtergrondkenmerken. Wij stellen voorop dat de volgende analyses geenszins pretenderen een uitputtend inzicht te geven in mogelijke verklaringen voor begrip of gebrek aan begrip van de conclusies. Eerder zouden de analyses gezien moeten worden als evaluatie van enkele voor de hand liggende factoren die aan begrip zouden kunnen bijdragen.

Omdat rechters en advocaten in vorenstaande analyses van begrip in statistische zin beschouwd kunnen worden als één homogene groep (dus geen significante verschillen in begrip), hebben we voor de volgende analyses beide groepen samengenomen om als zodanig te onderscheiden van de groep deskundigen van het NFI. Voor beide groepen worden afzonderlijk de analyses gepresenteerd. Dit komt doordat de achtergrondvragen in de vragenlijst voor deskundigen van wat andere aard waren dan voor de juristen.

### *Juristen (advocaten en rechters)*

Aan de juristen is tegen het eind van de vragenlijst een serie algemene vragen gesteld. Het betrof de volgende vragen (nadruk alhier toegevoegd). Het percentage positieve antwoorden staat er telkens bij.

Heeft u in de afgelopen jaren een **cursus gevolgd** over onderwerpen zoals de wijze waarop deskundigen hun conclusies formuleren en de interpretatie daarvan in de rechtszaal? (45%)

Leest u met enige regelmaat **vakliteratuur of wetenschappelijke literatuur** over deze onderwerpen? (47%)

Heeft u de **vakbijlage van het NFI** over dit onderwerp gelezen ('De reeks waarschijnlijkheidstermen van het NFI en het Bayesiaanse model voor de interpretatie van bewijs')? (12%)

In technisch-forensische rapporten worden conclusies in toenemende mate geformuleerd in termen van de relatieve waarschijnlijkheid van de bevindingen onder twee hypothesen of scenario's. Bent u deze wijze van concluderen in concrete strafzaken **al tegengekomen**? (41%)

Hoe **vertrouwd** voelt u zichzelf als het gaat om het lezen en begrijpen van cijfermateriaal en statistische gegevens? (4,2 gemiddeld, op een schaal van 1 tot 7).

Ten slotte werd gevraagd naar **leeftijd** (48 jaar gemiddeld), **geslacht** (43% vrouw) en **ervaring** in het strafrecht (13 jaar gemiddeld).

In een multiële regressie-analyse werd onderzocht in welke mate feitelijk begrip kon worden verklaard met voornoemde informatie. Het resultaat van deze analyse is dat de gevraagde achtergrondinformatie in het geheel niet bijdraagt aan een verklaring van feitelijk begrip als optelsom van de begripstellingen in de twee casus.<sup>41</sup>

Met behulp van logistische regressie-analyse werd ten slotte ingezoomd op de *prosecutor's fallacy* in beide rapporten. Zowel bij de overval op het tankstation als bij de beroving op straat was aan de hand van logistische regressie met genoemde achtergrondvariabelen geenszins te voorspellen welke personen de stelling zouden herkennen als fout.<sup>42</sup>

## Deskundigen

Aan de deskundigen van het NFI werden de volgende vragen gesteld.

Heeft u in de afgelopen jaren een **cursus** gevolgd waarin het bayesiaanse denkkader voor interpretatie van bewijs werd uitgelegd (bijvoorbeeld de NFI-cursus 'Bayes' en Criminalistiek)? (85%)

Leest u met enige regelmaat **vakliteratuur of wetenschappelijke** literatuur over deze onderwerpen? (51%)

Heeft u de **vakbijlage** van het NFI over dit onderwerp gelezen ('De reeks waarschijnlijkheidstermen van het NFI en het Bayesiaanse model voor de interpretatie van bewijs')? (80%)

Hoe **vertrouwd** voelt u zichzelf als het gaat om het lezen en begrijpen van cijfermateriaal en statistische gegevens? (5,3 gemiddeld)

<sup>41</sup> Het regressiemodel met alle variabelen daarin geforceerd opgenomen, resulteerde in een totale verklaarde variantie  $R^2$  van 3%.

<sup>42</sup> Logistische regressie wordt gebruikt wanneer de te verklaren variabele een ja/nee-variabele is. In ons geval: is de stelling correct geclassificeerd, ja of nee? De logistische regressievariant van de verklaarde variantie is Nagelkerke's  $R^2$ . Voor de separate logistische analyses bij de tankcasus en de straatcasus was deze respectievelijk 6% en 4%.

Bent u **tekenbevoegd**? (72%)

Gebruikt u bij het formuleren van conclusies **in uw rapporten de systematiek van 'logisch correct redeneren'** (d.w.z. conclusie in de vorm van een (verbale) *likelihood ratio*)? (36%)

Ook aan de deskundigen werd gevraagd naar **leeftijd** (40 jaar gemiddeld) en naar **geslacht** (38% vrouw).

De multiële regressie-analyse bij de deskundigen leverde een resultaat op dat naar de meeste maatstaven mager te noemen is. Met alle variabelen opgenomen in het regressiemodel kon 17% ( $R^2$ ) van de variantie in feitelijk begrip onder deskundigen worden verklaard. De enige variabele met een gewicht van noemenswaardige omvang in de regressievergelijking was het antwoord op de vraag of men zelf concludeert in 'logisch correcte' vorm.<sup>43</sup>

Ook bij de deskundigen is aan de hand van logistische regressies getracht het begaan van de *prosecutor's fallacy* in beide casus te verklaren met genoemde achtergrondvariabelen. Bij de conclusie in de casus over de overval op het tankstation resulteerde dat in een verklaarde variantie van 23%.<sup>44</sup> Voor de correcte classificatie van de *prosecutor's fallacy* in het kader van die casus bleek het gevolgd hebben van de cursus en in iets mindere mate het zelf in deze vorm rapporteren van significante invloed. Van de deskundigen die de interne cursus over deze vorm van rapporteren hebben gevolgd, herkende 45% de *prosecutor's fallacy* bij dit rapport. Van de deskundigen die die cursus niet hebben gevolgd, herkende 27% de fout. Merk op dat hoewel de cursus wel degelijk effect sorteert, de meerderheid ook na het volgen van die cursus nog de fout maakt. De deskundigen die zelf op deze wijze rapporteren, bewijzen dat de praktijk voor henzelf de beste leerschool is. Van hen herkent 63% de *prosecutor's fallacy* in deze zaak, terwijl van degenen die een andere conclusievorm hanteren 31% de valkuil weet te vermijden. Ten slotte, van de deskundigen die zelf op deze wijze rapporteren én de cursus hebben gevolgd ( $N = 28$ ), herkent 64% de *prosecutor's fallacy*, terwijl slechts één van de tien deskundigen die noch de cursus hebben gevolgd, noch zelf zo rapporteren deze denkfout herkent.

Bij de casus over de beroving op straat lag de verklaarde variantie op een wat hoger niveau (27%) en bleek enkel het gevolgd hebben van de cursus in statistisch significante zin een bijdrage te leveren. Van degenen die de cursus hebben gevolgd, herkende 43% de *prosecutor's fallacy* als fout. Slechts 7% van deskundigen die niet in de schoolbank bij de NFI-afdeling 'statistiek' hadden gezeten, weet om deze valkuil heen te lopen, wanneer zij de conclusie over de

43 Gestandaardiseerde regressiecoëfficiënt 0,26;  $t = 2,27$ ,  $p = 0,03$ .

44 Nagelkerke's  $R^2$ . Zie verder voor details van de analyse tabel A2.12.

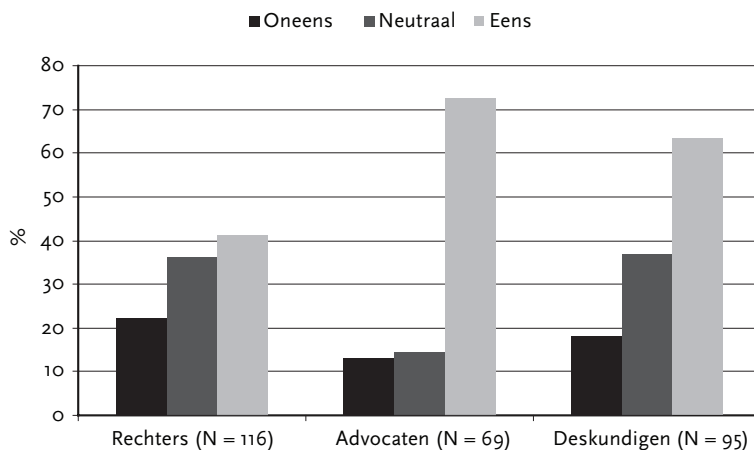
beeldvergelijking bij de beroving op straat bestuderen. Van de deskundigen die de cursus hebben gevolgd en zelf deze methodiek van rapporteren toepassen (N = 28), ziet 61% de *prosecutor's fallacy*, en niet één van de tien deskundigen die noch de cursus hebben gevolgd, noch zelf zo rapporteren, herkent deze denkfout.

## ENKELE BESCHOUWINGEN OVER 'HET PROBLEEM'

Aan de deelnemers aan Studie 2 legden wij ter afsluiting enkele algemeen beschouwende stellingen voor over de problematiek van communicatie tussen deskundigen en de juridische professionele procesdeelnemers.

### Wiens probleem?

Figuur 10 geeft de reacties weer op de stelling: *Als deskundigenrapporten niet goed begrepen worden, is dat een probleem dat de deskundige dient op te lossen*. Respondenten gaven hun mening op een 7-puntsschaal lopend van 1 ('helemaal mee oneens') tot 7 ('helemaal mee eens'). In de figuur zijn de antwoorden in drie categorieën verdeeld, waarbij de categorie 'neutraal' het schaal midden, score 4, betreft.



Figuur 10 Reacties op stelling 'Onbegrip is probleem dat deskundige dient op te lossen', Studie 2

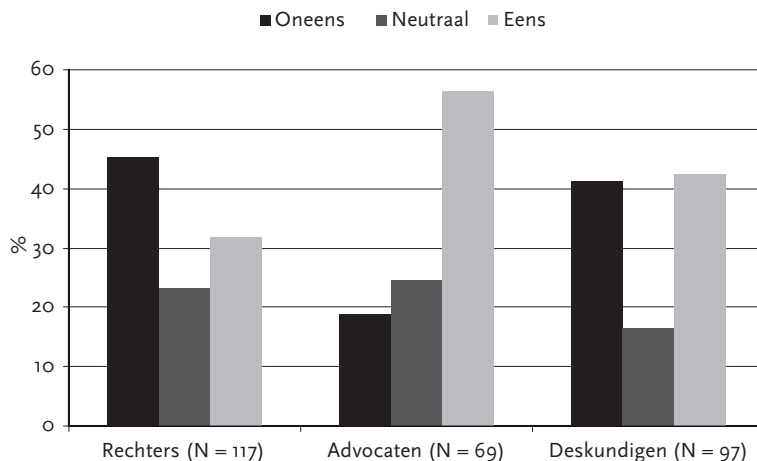
Een drietal zaken valt op in figuur 10. Ten eerste valt te constateren dat de deskundigen zelf zich niet verschuilen. Ruim 63% van de deskundigen is het met deze stelling eens. Bij de rechters is het aantal instemmers een stuk lager (41%). Interessant is dat ruim 22% van de rechters het er niet mee eens is dat de deskundigen het maar moeten oplossen. Voorzichtig stellen wij dat

de rechters zich het probleem kennelijk ook zelf aanrekenen. Voor de advocaten, ten slotte, lijkt een en ander klip en klaar: driekwart (73%) meent dat de deskundigen verantwoordelijk zijn voor oplossing van misverstanden over hun rapporten.

### **Defaitisme of realisme?**

Ook werd de nogal defaitistische (of moeten we inmiddels zeggen: realistische?) stelling voorgelegd: *Het grote verschil in vakspecifieke kennis en referentiekader tussen technisch deskundige en jurist zorgt ervoor dat iedere dialoog tussen beide altijd gekenmerkt wordt door kloven en misverstanden.*

Figuur 11 vat (op dezelfde wijze als de vorige figuur) de reacties op deze stelling samen. Figuur 11 laat zien dat een dergelijke opstelling de meerderheid van de rechters niet eigen is. Weliswaar stemt ruim 30% van hen in met deze stelling, een hoopvolle 45% van de rechters is het er simpelweg niet mee eens. Voor de meeste advocaten lijken de kloven en misverstanden een 'fact of life': 57% stemt in. Onder de deskundigen is sprake van wat wij geneigd zijn een schisma te noemen. Is ruim 42% van hen een positieve attitude aangedaan over de kloof tussen jurist en deskundige, een groep van even grote omvang schijnt de moed te hebben opgegeven.



**Figuur 11** Reacties op stelling: 'Het grote verschil in vakspecifieke kennis en referentiekader tussen technisch deskundige en jurist zorgt ervoor dat iedere dialoog tussen beiden altijd gekenmerkt wordt door kloven en misverstanden', Studie 2

### ***Over kennis en begrip***

Aan de rechters en aan de advocaten is gevraagd of zij menen dat binnen de eigen beroepsgroep voldoende specifieke kennis aanwezig is om goed te begrijpen wat deskundigen in hun rapportages tot uitdrukking trachten te brengen. Terwijl slechts weinig rechters en advocaten over hun eigen vakbroeders meldden dat men zonder meer tekortschiet (resp. 8% en 16%), was het antwoord 'ja zonder meer' nochtans even weinig of minder populair (resp. 9% en 3%). Verreweg de meeste reacties ten aanzien van de eigen beroepsgroepen betroffen de antwoorden, *Ja, maar het houdt niet over* (rechters 45%, advocaten 32%) en *Nee, men schiet enigszins tekort* (resp. 39% en 49%). Gegeven de zojuist beschreven bevindingen van Studie 2, komen deze oordelen ons inmiddels toch ongerechtvaardigd mild voor.

Een wat realistischer reflectie op het begrip van 'logisch correct' geformuleerde conclusies onder rechters en advocaten vonden wij bij de deskundigen. Gevraagd of zij de indruk hebben dat rechters, officieren van justitie en advocaten conclusies volgens de systematiek van 'logisch correct redeneren' op de juiste wijze weten te interpreteren, gaf 86% als antwoord: 'Nee'.



## Conclusies

De communicatie tussen technisch-forensische deskundigen en professionele procesdeelnemers in de rechtszaal is een onderwerp dat reeds lange tijd in de belangstelling staat. Recente spraakmakende strafzaken, waarbij deskundigenonderzoek en -rapportage een rol speelden, hebben die aandacht alleen maar doen toenemen. Enerzijds is de maatschappelijke aandacht voor het werk van de deskundige en diens rol in strafzaken gegroeid, anderzijds zijn ook de professionele partijen in toenemende mate het belang gaan inzien van goed begrip van en communicatie over het onderzoek van de deskundigen. Bij het Nederlands Forensisch Instituut heeft een en ander geleid tot twee zichtbare ontwikkelingen.

De eerste ontwikkeling is de uniformering en professionalisering van het geschreven eindproduct van technisch-forensisch onderzoek. De eisen die daaraan gesteld worden, zijn steeds strenger geworden, niet alleen met het oog op de wetenschappelijke inhoud, maar ook met het oog op de presentatie van onderzoek en bevindingen naar de uiteindelijke consument van de rapportage. In dat licht moet ook de vlucht worden gezien die de vakbijlagen hebben genomen die het NFI meestuurt met individuele onderzoeksrapportages. In die vakbijlagen tracht het NFI concreet onderzoek, methoden en beperkingen daarvan in voor leken begrijpelijke taal en vorm uit te leggen om op die wijze het begrip van zaaksgerelateerde rapportages te waarborgen. Tegenwoordig beschikken de meeste afdelingen op het NFI over vakbijlagen die er uitermate professioneel uitzien en doelbewust in samenspraak met niet-deskundigen zijn en worden ontwikkeld. Een goed voorbeeld daarvan is de uitgave 'Essenties van forensisch DNA-onderzoek' (Meulenbroek, 2008).

De tweede ontwikkeling is de ingebruikname van de methodiek van 'logisch correct rapporteren', op de wijze zoals hiervoor uitvoerig is besproken. Door onderzoeksbevindingen niet langer in een herkomstuitspraak te gieten (zoals voorheen veelal gebeurde), maar in de vorm van een aannemelijkheidsverhouding, een diagnostische waarde, te presenteren, treedt de deskundige niet (langer) buiten zijn of haar deskundigheid. Wat er vervolgens met dat stukje bewijs gebeurt in een rechtszaak, dat is aan de rechter. In 'logisch correcte' zin kan de deskundige niet meer dan het uitdrukken van die diagnostische waarde van zijn bevindingen. Al wat daaraan voorbijgaat, en dat gebeurde voorheen vaak, is in principe logisch incorrect. Op basis van een waargenomen overeenkomst tussen een betwist stuk van overtuiging en referentiemateriaal van een verdachte, is geen zelfstandige uitspraak te doen over de kans dat verdachte de donor is van het betwiste stuk.



Er is inmiddels een wat paradoxale situatie ontstaan. Vanuit de rechtspraktijk klinken nu weemoedige geluiden dat de vroegere wijze van rapporteren veel duidelijker was en de voorkeur verdient. Waarom toch zo moeilijk doen met hypothesen en scenario's? Welnu, praktijkjuristen kunnen dan wel menen dat het vroeger allemaal beter en duidelijker was, feit is dat die wijze van rapporteren incorrect was. Er is geen weg meer terug! En strikt genomen is dat maar goed ook.

Het grote probleem, en dat was de kern van ons onderzoek alhier, is dat begrip en communicatie op dit terrein kennelijk achterlopen bij voortschrijdend forensisch wetenschappelijk inzicht. Door ons is de omvang van een eventueel probleem in kaart gebracht door te onderzoeken hoe goed of slecht professionele procesdeelnemers forensisch-technische conclusies in de vorm van aannemelijkheidsverhoudingen begrijpen. Een van de ideeën die ten grondslag aan het onderzoek lag, was dat een eventueel tekortschietend begrip in de praktijk een ernstiger probleem zal zijn naarmate de gebruikers van de rapporten zich minder bewust zijn van hun eigen onbegrip: feitelijk begrip is immers niet hetzelfde als vermeend begrip.

In de voorstudie met studenten werd duidelijk dat bij zulk hoog opgeleide personen, het feitelijk begrip van forensisch-technische conclusies in de vorm van verbale aannemelijkheidsverhoudingen tekortschiet. Tegen de verwachting in, maakte het daarbij voor feitelijk begrip niet uit of men een stevige wiskundige achtergrond had of geheel niet. Het ging dan om spontaan begrip van de conclusies, dus zonder voorafgaande uitleg van de conclusievorm. De grote meerderheid van de deelnemers aan de voorstudie was voorts van mening dat ze de conclusies in de hen voorgelegde rapporten goed tot zeer goed had begrepen. Dat hoge vermeend begrip werd gelogenstraft door het feitelijk begrip bij diezelfde personen. Er was bij de studenten zonder meer sprake van een forse kennisoverschatting.

Naar aanleiding van de ervaringen in de voorstudie met studenten kreeg de hoofdstudie definitieve vorm. In deze studie werden grofweg dezelfde methode en procedure gevolgd: het voorleggen van fictieve deskundigenrapporten met daarin de conclusies in de vorm van aannemelijkheidsverhoudingen. Feitelijk begrip werd gemeten aan de hand van stellingen verwoord als goede en foute interpretatie van de conclusies. Vermeend begrip werd gemeten door de respondent direct te vragen hoe goed hij of zij de conclusies had begrepen. Vooruitlopend (en met de resultaten van de voorstudie vers in het geheugen) op tekortschietend begrip van de verbale aannemelijkheidsverhoudingen, hebben we visuele varianten van dezelfde conclusies in het onderzoek betrokken. Die varianten voldeden aan de eisen van logische correctheid, maar drukten de aannemelijkheidsverhouding non-verbaal uit. Wellicht, zo was onze gedachte, zou dit een aanwijzing opleveren in een richting waarin verbetering van begrip mogelijk is.

Drie beroepsgroepen participeerden in de hoofdstudie: rechters en raadsheeren die werkzaam zijn in de strafsectoren van een zevental arrondissementsrechtbanken en gerechtshoven, strafrechtadvocaten en medewerkers van het NFI. Van die laatste groep is 72% tekenbevoegd deskundige en 28% in opleiding tot deskundige. Van de NFI-mensen meldt 36% de methode van logisch correct rapporteren thans te gebruiken. De deskundigen figureerden als onderzoeksgroep in dit project vanuit de gedachte dat we ook een goed beeld zouden moeten hebben van feitelijk en vermeend begrip van juist degenen die begrip nu en in de toekomst bij anderen zouden moeten bewerkstelligen. Dat het bevragen van deze specifieke groep een verstandige keuze was, blijkt klip en klaar uit de onderzoeksresultaten.

Ten opzichte van de vroegere wijze van rapporteren, zou de logisch correcte methode op twee fronten een vooruitgang moeten bewerkstelligen. In de eerste plaats zit in de oude methode de *prosecutor's fallacy* ingebakken, terwijl de logisch correcte methode de interpretatie in termen van de waarschijnlijkheid van de bevindingen gegeven de hypothese aanreikt. Toch blijkt de *prosecutor's fallacy* ook nu nog in een ruime meerderheid van de gevallen gemaakt te worden, uiteenlopend van meer dan 50% bij de medewerkers van het NFI tot meer dan 90% bij de rechters. Weliswaar is dat minder dan 100% zoals bij de oude methode het geval was, maar de winst is nu nog tamelijk gering.

In de tweede plaats drukt rapportage in termen van een aannemelijkheidsverhouding uit dat men de bevindingen in het licht van twee hypothesen dient te interpreteren. Inderdaad blijken rechters, advocaten en deskundigen in grote meerderheid (circa 90%) die gedachte te herkennen en te onderschrijven, mits rechtstreeks in deze termen geformuleerd. Maar we moeten ook concluderen dat de respondenten niet 'boven de stof' staan, want veel minder positief zijn de uitslagen als we parafrases van juiste of onjuiste conclusies voorleggen. Totaal genomen, over alle vragen, weet minder dan één op de vijf rechters, minder dan één op de tien advocaten en minder dan één op de twee NFI-medewerkers vrijwel foutloos te antwoorden. Kortom, het onderzoek laat zien dat rechters en advocaten conclusies in deskundigenrapportages nog niet goed begrijpen wanneer deze in de logisch correcte modus, met behulp van een aannemelijkheidsverhouding geformuleerd zijn. Daarbij doet het er niet toe of die verhouding visueel of verbaal is uitgedrukt. Kennelijk is het concept van de aannemelijkheidsverhouding en de strekking daarvan voor de meeste personen moeilijk te begrijpen (hoe logisch correct deze ook mag zijn). Deze bevinding hebben wij gekwalificeerd als zorgelijk. De reden daarvoor is dat in de dagelijkse rechtspraktijk met deze rapportages wordt gewerkt en er mede op basis van dergelijke rapporten beslissingen worden genomen die ingrijpend zijn voor echte mensen. Natuurlijk weten wij niet of en in welke mate dat dan in de rechtspraktijk leidt tot foute beslissingen. Wij constateren hier slechts

dat een belangrijke bouwsteen van rechterlijke beslissingen door rechters (en door anderen) niet goed begrepen wordt, en dat noemen wij zorgelijk.

De kwalificatie 'zorgelijk' zou te nuanceren zijn door te wijzen op het enigszins artificiële karakter van een onderzoek als het onderhavige. Zo hebben de deelnemers in ons onderzoek zonder te kunnen overleggen een compacte casus en de bijbehorende schriftelijke vragenlijst beoordeeld. In de praktijk echter kan de rechter overleggen met collega's en zich op zitting laten voorlichten door de deskundige die het rapport schreef. De denkfouten bij interpretatie van de conclusies zijn volgens onze studie echter zo wijdverbreid bij procesdeelnemers, dat misverstanden mogelijk eerder hardnekkiger worden dan dat ze worden opgelost. Op de verhelderende rol die de deskundige hierin zou kunnen spelen, komen we zo dadelijk terug.

Eerst betogen we dat het probleem nog lastiger is dan op het eerste gezicht lijkt. Er is namelijk sprake van een flinke kennisoverschatting bij rechters en advocaten. Terwijl deze professionele procesdeelnemers menen de conclusies goed te begrijpen, wordt feitelijk begrip bij diezelfde personen overwegend gekenmerkt door misverstanden, zoals de *prosecutor's fallacy* en de *defense fallacy*. Dit maakt het probleem lastiger, doordat mensen die niet weten dat ze fouten maken, dat naar alle waarschijnlijkheid zo zullen blijven doen en geen directe noodzaak percipiëren tot verbetering. Wat dat betreft, is te hopen dat ons onderzoek tenminste aan het besef van de feitelijke situatie bijdraagt en als zodanig vermeend begrip helpt terug te brengen naar het niveau van het tekortschietend feitelijk begrip: het niveau waar het op dit moment thuishoort. Overigens willen wij hier nog eens opmerken dat vanwege de lage respons onder de advocaten, met het generaliseren van de onderzoeksresultaten naar die beroepsgroep enige voorzichtigheid op haar plaats is.

Van deskundigen die logisch correct rapporteren, zou mogen worden verwacht dat zij volledig boven deze materie staan en de notoire valkuilen weten te ontwijken. Voor zover dat niet het geval is, zou toch verwacht kunnen worden dat die deskundigen zich dan bewust zijn van een tekortschietend begrip van (verbale) *likelihood ratio's* in forensisch-technische conclusies. Het onderzoek liet zien dat deskundigen en deskundigen in opleiding veel beter presteren dan rechters en advocaten. Het feitelijk begrip bij de mensen van het NFI is beduidend groter dan bij de juristen. Toch worden ook door NFI-medewerkers zelf veel fouten gemaakt. Juist denkfouten waar ook binnen het NFI zo veel aandacht voor is, zoals de *prosecutor's fallacy*, worden op grote schaal gemaakt. Op basis van de hier gepresenteerde onderzoeksresultaten kan niet geconcludeerd worden dat die mensen in voldoende mate boven de materie staan om feitelijk onbegrip bij rechters en advocaten te signaleren en adequaat te corrigeren. Daarbij merken we nogmaals op dat thans slechts 36% van de mensen van het NFI in onze studie deze logisch correcte methode gebruikt bij het rapporteren. Ook bij de medewerkers van het NFI signaleren we

een kennisoverschatting. De kloof tussen vermeend begrip en feitelijk begrip is weliswaar niet zo groot als bij de juristen, toch is er wel degelijk sprake van zo'n kloof.

In een poging het feitelijk begrip te verklaren, kan het volgende geconcludeerd worden. Bij de rechters en bij de advocaten kon aan de hand van enkele voor de hand liggende achtergrondgegevens geen enkele bijdrage worden geleverd aan de verklaring van begrip. Naast wat standaardgegevens als leeftijd en geslacht, bleek ook het hebben gevolgd van algemene instructies en het lezen van vakliteratuur over dit onderwerp van weinig waarde voor beter begrip. Kennelijk is de kern van de methode van concluderen zelf zo lastig te begrijpen, dat bestaande algemene instructies geen effect sorteren. Dat is natuurlijk in lijn met de bevinding uit Studie 1, dat het voor spontaan feitelijk begrip niet uitmaakte of men wiskunde studeert of rechten. Bij de mensen van het NFI was het daarentegen wel mogelijk in beperkte mate variatie in begrip te verklaren. De twee kenmerken die daarin een rol van betekenis speelden, waren of de desbetreffende deskundige zelf van deze methode gebruikmaakt bij het schrijven van rapporten en of men de interne cursus over logisch correct rapporteren heeft gevolgd.

Dat deskundigen en deskundigen in opleiding zich er, ondanks de eigen kennisoverschatting, van bewust zijn dat er problemen zijn in de communicatie tussen juristen en deskundigen, blijkt desgevraagd uit antwoorden op de algemeen beschouwende vragen die we hun stelden. De meesten vonden dat eventueel onbegrip over rapporten van het NFI een probleem is dat de deskundigen zelf moeten oplossen. Tegelijkertijd signaleerden we een schisma bij de medewerkers van het NFI, waarbij de ene groep, van substantiële omvang, de handdoek in de ring scheen te gooien, en de andere, ongeveer even grote groep, van mening was dat een en ander wel op te lossen zou zijn. Overigens hadden de NFI-medewerkers toch een behoorlijk adequaat beeld van het feitelijk onbegrip bij juristen in de rechtszaal. Op de vraag of rechters, officieren van justitie en advocaten de conclusies volgens de systematiek van 'logisch correct redeneren' op de juiste wijze weten te interpreteren, gaven bijna alle NFI-medewerkers het antwoord 'Nee'.



## Hoe nu verder?

Op grond van de resultaten van deze studie kunnen we ons afvragen welke gevolgtrekkingen aan de bevindingen dienen te worden verbonden. We verlaten daarmee het rapporteren van het empirisch onderzoek als zodanig en formuleren wat wij, de auteurs van dit rapport, menen dat er op grond van onze bevindingen kan of moet worden gedaan.

### COMMUNICATIEPROBLEEM

Ondubbelzinnig is aangetoond dat, hoezeer de methode van de aannemelijkheidsverhouding ook logisch correct is, het nog niet lukt om dat wat de deskundige wil en behoort over te brengen ook voldoende over te brengen. De kern van het probleem zit hem in de communicatie tussen makers en gebruikers van deze deskundigenrapportages. Het gegeven dat professionele procesdeelnemers een belangrijke bouwsteen voor hun beslissing en overtuiging niet begrijpen, vinden wij inherent problematisch. Onzes inziens kunnen we als rechtsgemeenschap deze stand van zaken niet accepteren. Het is waar dat veruit de meeste strafzaken zonder zulke forensische deskundigenrapportages worden beslist. Het is ook juist dat we niet weten hoe vaak in zaken waar zulke rapportages wel aan de orde zijn onvolledig begrip tot verkeerde uitspraken van rechters leidt; daarover doet dit onderzoek geen uitspraak. Maar dat door verbetering van forensische conclusies en het begrip daarvan het risico op zulke fouten vermindert, staat buiten kijf, en dat lijkt ons daarom een desideratum dat hoog op de agenda moet staan. Bij het NFI staat het daar thans ook. Daarbij ligt het voor de hand als eerste de *prosecutor's fallacy*, als de meest hardnekkige, zij het niet enige, denkfout aan te pakken.

### GEEN WEG TERUG

Wat kan er dan worden gedaan? We stellen de diagnose dat de vroegere vorm van rapporteren incorrect was niet ter discussie. Er is dus geen weg terug, de nostalgische verzuchting van sommige procesdeelnemers ten spijt. Er zal moeten worden gewerkt aan het bewerkstelligen van beter begrip binnen de kaders van het nieuwe paradigma. Een eerste en belangrijke stap daartoe is met dit onderzoek reeds genomen. Voordat feitelijk begrip kan worden verbeterd, dient men er ten minste van doordrongen te zijn dat het eigen vermeend begrip een misvatting was. Pas dan staat de cognitieve deur open voor een

positief leerproces. Het bestaan en de omvang van de kennisoverschatting hopen wij hiervoor overtuigend te hebben aangetoond.

## WAT IS DE OPLOSSING NIET?

In het kader van dit onderzoek is ook een oplossingsrichting verkend door te experimenteren met een visueel uitgedrukte aannemelijkheidsverhouding. Dit verhoogde het begrip niet. Het is kennelijk niet zozeer de *vorm* waarin de aannemelijkheidsverhouding wordt gepresenteerd, maar de complexiteit van de *essentie* van wat een aannemelijkheidsverhouding precies tracht over te brengen, die vaak tot onbegrip leidt. De oplossing moet dus niet primair in de 'cosmetische' hoek worden gezocht. Voor medewerkers van het NFI bleek dat opleiding en vooral ervaring wel bijdragen aan begrip. Men zou dus kunnen investeren in meer, specifiekere en intensievere cursussen, ook voor anderen dan deskundigen. Toch zijn wij wat dat betreft niet erg optimistisch. Immers, de cursussen zoals het NFI die thans aan zijn medewerkers aanbiedt, blijken hen op slechts beperkte schaal te beschermen tegen valkuilen als de *prosecutor's fallacy*. Ook het gegeven dat voor studenten wiskundige aanleg, vaardigheid of opleiding niet echt bijdraagt aan spontaan feitelijk begrip, geeft te denken. Het is dus ook niet in eerste instantie een probleem van het verbeteren van de algemene kwantitatieve vaardigheid of leren omgaan met wiskundige en probabilistische begrippen. Het is een probleem gelegen in begrip van de essentie van de methode. De indruk dat het probleem niet alleen bij de individuele rechter, advocaat, officier of deskundige ligt, maar algemeen is, wordt versterkt doordat de problemen die wij nu voor Nederland hebben onderzocht zich, volgens experts, in andere landen eveneens voordoen (Taroni & Aitken, 1998).

## STUDIEGROEP NODIG

Alle partijen zouden meer moeten investeren in grip krijgen op de precieze strekking van aannemelijkheidsverhoudingen en hoe die moet worden overgebracht, om daar vervolgens *open-minded* met elkaar over van gedachten te wisselen. Voor een belangrijk deel vraagt dat om diepgaander opleiding, maar dan wel andersoortig dan bestaande cursussen. Opleiding moet zich wellicht niet zozeer op de kwantitatieve techniek richten, maar allereerst en vooral op conceptuele verheldering: wat is de rol van de geboden argumentatie precies bij de taak van de rechter? Het lijkt er nu op dat er impliciet of expliciet geen consensus bestaat tussen de uiteenlopende partijen die met dergelijke rapporten te maken krijgen over de interpretatie van aannemelijkheidsverhoudingen. Het lijkt zaak een krachtige poging te doen om hierin verheldering te verkrijgen en vooral om veel betere methoden te ontwikkelen om dat effectief

en beklijvend te communiceren. Is het niet tijd voor een fundamentele bezinning, door een gemengde groep van strafrechters, officieren, advocaten, straf(proces)rechtsgeleerden, forensisch deskundigen, probabilisten, rechtsfilosofen, rechtspsychologen en communicatiedeskundigen, om de kern van het communicatieprobleem in kaart te brengen? Wat moet er worden overgebracht? En vooral ook: hoe doen we dat? Een onberispelijke strafrechtspleging is te belangrijk om dit probleem te laten liggen.

Wij willen wel een voorschot nemen op onze inbreng in zo'n discussie. Wij menen dat heersend onbegrip deels is gerelateerd aan het als *package deal* aanbieden van uitleg over aannemelijkheidsverhoudingen en de interpretatie ervan in een bayesiaans kader. Sterker geformuleerd: de onmiddellijke integratie van aannemelijkheidsverhoudingen binnen de regel van Bayes als prescriptief model van juridisch redeneren, staat voor veel mensen (juristen en deskundigen) het essentiële begrip van de aannemelijkheidsverhouding in de weg. Het standpunt dat men een aannemelijkheidsverhouding als middel voor bewijsevaluatie toejuicht en toch geen bayesiaan is, komt ons alleszins verdedigbaar voor.





## Literatuur

- Broeders, A.P.A. (1999). Some observations on the use of probability scales in forensic identification. *Forensic Linguistics*, 6, 228-241.
- Broeders, A.P.A. (2005). Individualisatie in de traditionele criminalistiek. In: M.J. Sjerps & J.A. Coster van Voorhout (red.). *Het onzekere bewijs. Gebruik van statistiek en kansrekening in het strafrecht* (pp. 49-97). Deventer: Kluwer.
- Brun, W. & K.H. Teigen (1988). Verbal Probabilities: Ambiguous, Context-Dependent, or Both? *Organizational Behavior and Human Decision Processes*, 41, 390-404.
- Champod, C. & I.W. Evett (2000). Commentaries on Broeders' 'Some observations on the use of probability scales in forensic identification'. *Forensic Linguistics*, 7, 238-243.
- Crombag, H., P.J. van Koppen & W.A. Wagenaar (2002). De waarde van het bewijs. In: P.J. van Koppen, D.J. Hessing, H. Crombag & H. Merckelbach (eds.). *Het recht van binnen: Psychologie van het recht* (pp. 281-293). Deventer: Kluwer.
- Evett, I.W. (1987). Bayesian inference and forensic science: problems and perspectives. *The Statistician*, 36, 99-105.
- Evett, I.W. (1995). Avoiding the transposed conditional. *Science & Justice*, 35, 127-131.
- Gatowski, S.I., S.A. Dobbin, J.T. Richardson, G.P. Ginsburg, M.L. Merlino & V. Dahir (2001). Asking the gatekeepers: A national survey of judges on judging expert evidence in a post-Daubert World. *Law and Human Behavior*, 25, 433-458.
- Grunberg, S.M., S. Groshen, S. Steingass, S. Zaretsky & B. Meyerowitz (1996). Comparison of conditional quality of life terminology and visual analogue scale measurements. *Quality of Life Research*, 5, 65-72.
- Kaasa, S.O., T. Peterson, E.K. Morris & W.A. Thompson (2007). Statistical Inference and Forensic Evidence: Evaluating a Bullet Lead match. *Law & Human Behavior*, 31, 433-447.
- Kerkmeester, H.O. (2005). Het gebruik van Bayesiaanse statistiek in strafprocessen. In: M.J. Sjerps & J.A. Coster van Voorhout (red.). *Het onzekere bewijs. Gebruik van statistiek en kansrekening in het strafrecht* (pp. 99-129). Deventer: Kluwer.
- Meulenbroek, A.J. (2008). *De essenties van forensisch DNA-onderzoek*. Den Haag: Nederlands Forensisch Instituut.
- Murphy, A.H., S. Lichtenstein, B. Fischhoff & R.L. Winkler (1980). Misinterpretations of Precipitation Probability Forecasts. *Bulletin American Meteorological Society*, 61, 695-701.
- NFI (2008). *Vakbijlage: De reeks waarschijnlijkheidstermen van het NFI en het Bayesiaanse model voor interpretatie van bewijs*. [www.forensischinstituut.nl/Images/NFI%20Vakbijlage%20Gezichtsvergelijking\\_tcm68-146785.pdf](http://www.forensischinstituut.nl/Images/NFI%20Vakbijlage%20Gezichtsvergelijking_tcm68-146785.pdf).

- Olson, M.J. & D.V. Budescu (1997). Patterns of Preference for Numerical and Verbal Probabilities. *Journal of Behavioral Decision making*, 10, 117-131.
- Robertson, B. & G.A. Vignaux (1995). *Investigating Evidence: Evaluating Forensic Science in the Courtroom*. Chicester: John Wiley & Sons.
- Sjerps, M.J. (2008). Forensische statistiek en kansrekening: Interpretatie van bewijs. In: A.P.A. Broeders & E.R. Muller (red.). *Forensische Wetenschap* (pp. 467-496). Deventer: Kluwer.
- Sjerps, M.J. (2000). Pros and cons of Bayesian reasoning in forensic science. In: J.F. Nijboer & W.J.J.M. Sprangers (eds.). *Harmonization in Forensic expertise* (pp. 557-585). Series Criminal Sciences, Amsterdam: Thela Thesis.
- Sjerps, M.J. & D.B. Biesheuvel (1999). The interpretation of conventional and 'Bayesian' verbal scales for expressing expert opinion: a small experiment among jurists. *Forensic Linguistics* 6, 214-227.
- Taroni, F. & C.G.G. Aitken (1998). Probabilistic reasoning and the law Part 1: assessment of probabilities and explanation of the value of DNA evidence. *Science & Justice*, 38, 165-177.
- Thompson, W.C. & E.L. Schumann (1987). Interpretation of statistical evidence in criminal trials: the prosecutor's fallacy and the defence attorney's fallacy. *Law and Human Behavior*, 11, 167-187.
- Thompson, W.C. (1989). Are juries competent to evaluate statistical evidence? *Law and Contemporary Problems*, 52, 9-40.

# Appendix 1 Onderzoeksmateriaal: deskundigenrapportages

Beroving op straat (*verbale* conclusievariant)  
Beroving tankstation (*visuele* conclusievariant)

## DESKUNDIGENRAPPORT

### Beroving op straat

*Onderwerp* Gezichtsvergelijking naar aanleiding van een beroving onder bedreiging van een mes bij een pinautomaat te Den Haag

### Onderzoeksmateriaal

*Ontvangen via* Technische Recherche Hollands Midden  
*Datum ontvangst* 28 december 2007

Code	Omschrijving
TR_A1	Cd met daarop digitaal beeldmateriaal; inhoud als volgend hieronder
SX1_B1	Foto frontaal aangezicht verdachte
SX1_B2	Foto profiel rechterzijde gezicht verdachte
SX1_B3	Foto profiel linkerzijde gezicht verdachte
SX1_C1	Stilstaand beeld van beveiligingscamera met onderschrift 'tijdcode 21:32:03'
SX1_C2	Stilstaand beeld van beveiligingscamera met onderschrift 'tijdcode 21:32:07'
SX1_C3	Stilstaand beeld van beveiligingscamera met onderschrift 'tijdcode 21:32:22'
SX1_C4	Stilstaand beeld van beveiligingscamera met onderschrift 'tijdcode 21:32:27'
SX1_C5	Stilstaand beeld van beveiligingscamera met onderschrift 'tijdcode 21:32:39'
SX1_C6	Stilstaand beeld van beveiligingscamera met onderschrift 'tijdcode 21:33:01'

#### *Toelichting op onderzoeksmateriaal*

De beelden SX1\_C1 t/m SX1\_C6 zijn afkomstig van de beveiligingscamera bij de pinautomaat. Verbalisant van de TR lichtte toe dat op ieder van die beelden de dader van de beroving voorkomt. Voor zover er twee personen te zien zijn, betreft het dader en slachtoffer (dit is het geval bij SX1\_C2, SX1\_C4, SX1\_C5). Dader is in deze beelden telkens te herkennen aan de baseballpet, terwijl slachtoffer geen hoofddeksel draagt.

### **Aanleiding onderzoek**

De volgende informatie is geleverd door de verbalisant.

Een jongeman die 's avonds op straat bij een pinautomaat zojuist € 100 heeft opgevraagd, wordt door een onbekende plotseling onder bedreiging van een mes dat bedrag afhandig gemaakt en gedwongen om direct nog eens € 500 te pinnen en af te geven. Naar aanleiding van een anonieme tip houdt de politie diezelfde avond een verdachte aan in een café. De technische recherche onderzoekt de volgende ochtend de opnames van de beveiligingscamera bij de pinautomaat en kopieert daarvan een zestal stilstaande beelden waarop de dader van de beroving zichtbaar is. Van verdachte worden foto's gemaakt op het politiebureau.

### **Vraagstelling**

*Is verdachte op de foto's SX1\_B1, SX1\_B2 en SX1\_B3 dezelfde persoon als de dader van de beroving op de beelden SX1\_C1 t/m SX1\_C6?*

Na een eerste screening van het aangeleverde beeldmateriaal op de cd (TR\_A1), werd duidelijk dat belichting, contrast en scherpte van de SX1\_C2 t/m SX1\_C6 van dusdanig slechte kwaliteit zijn, dat deze beelden ongeschikt zijn voor gezichtsvergelijking. SX1\_C1 is van matige kwaliteit, maar voldoende om in aanmerking te komen voor gezichtsvergelijking. Het betreft een beeld waarop de dader frontaal te zien is.

De vraagstelling is geïnterpreteerd als het verzoek om een vergelijkend onderzoek te verrichten en daarbij na te gaan of de bevindingen beter passen bij de hieronder weergegeven hypothese H1 dan bij de (alternatieve) hypothese H2.<sup>45</sup>

*Hypothese 1: De dader van de beroving zichtbaar op SX1\_C1 is dezelfde persoon als verdachte afgebeeld op foto SX1\_B1.*

*Hypothese 2: De dader van de beroving zichtbaar op SX1\_C1 is niet dezelfde persoon als verdachte afgebeeld op foto SX1\_B1.*

### **Onderzoek**

De gebruikte methode voor persoonsidentificatie aan de hand van gezichtsvergelijking op beeldmateriaal is gebaseerd op het vergelijken van morfologisch-antropologische kenmerken. Het aantal en de mate van gelijkheid van de gevonden kenmerken zijn de basis voor de gerapporteerde mate van steun

<sup>45</sup> De hypothesen zijn geformuleerd op grond van de gegeven informatie over de toedracht.

voor de hypothesen dat de te vergelijken afbeeldingen wel of niet dezelfde persoon betreffen. Met behulp van software is het beeldmateriaal eerst zo veel mogelijk geoptimaliseerd door aanpassing van contrast, helderheid en inzoomen op details.

### **Resultaten, interpretatie en conclusie**

Vergelijking van de gezichtskenmerken van de personen op foto SX1\_C1 en SX1\_B1, geeft een veelheid aan overeenkomsten in morfologische en antropologische kenmerken. Uit wetenschappelijk onderzoek is mij bekend dat deze combinatie van gelaatskenmerken in de Nederlandse bevolking zeldzaam is. Ik concludeer nu het volgende:

De bevindingen van de hier gerapporteerde gezichtsvergelijking aan de hand van het geselecteerde beeldmateriaal zijn *veel waarschijnlijker* wanneer de daarop afgebeelde persoon dezelfde persoon is (hypothese 1) dan wanneer het een andere persoon betreft (hypothese 2).<sup>46</sup>

<i>Ondertekening</i>	Aldus opgemaakt en getekend op de door mij afgelegde algemene eed/belofte als vast gerechtelijk deskundige
<i>Plaats</i>	Den Haag
<i>Datum</i>	11 januari 2008

<sup>46</sup> *Toelichting op de conclusie*

Voor het vergelijkend onderzoek wordt gebruikgemaakt van de volgende conclusiereeks waarmee een uitspraak wordt gedaan over de waarschijnlijkheid van de bevindingen, in het licht van de gestelde hypothesen. De gebruikte reeks is:

De bevindingen van het onderzoek zijn ...

- ongeveer even waarschijnlijk ... [als]
- iets waarschijnlijker ...
- waarschijnlijker ...
- veel waarschijnlijker ...
- zeer veel waarschijnlijker ...

... wanneer hypothese 1 juist is, dan wanneer hypothese 2 juist is.

## DESKUNDIGENRAPPORT

### Beroving tankstation

*Onderwerp* Chemische profilering van tape in relatie tot gewapende overval tankstation

#### **Onderzoeksmateriaal**

Ontvangen via KLPD  
Datum ontvangst 12 december 2007

Code	Omschrijving
SVO1.01	Stuk gebruikte blauwe tape gebruikt tijdens overval, lengte ca. 83 cm
SVO1.02	Stuk gebruikte blauwe tape gebruikt tijdens overval, lengte ca. 98 cm
SVO2.01	Rol blauwe tape uit woning verdachte

#### **Aanleiding onderzoek**

De volgende informatie is geleverd door de verbalisant.

Bij een gewapende overval op een tankstation langs de N13 heeft de gemaskeerde overvaller tape gebruikt om de in het tankstation aanwezige pompbediende te knevelen. Naar aanleiding van een anonieme tip is een dag na de overval een verdachte in zijn woning aangehouden. Bij zoeking in de woning vond de politie op het aanrecht een rol blauwe tape die in beslag is genomen voor forensisch onderzoek.

#### **Vraagstelling**

*Onderzoek de mate van overeenkomst tussen de tapes van SVO1.01/ SVO1.02 met de rol tape van SVO2.01.*

De vraagstelling is geïnterpreteerd als het verzoek om een vergelijkend onderzoek te verrichten en na te gaan of de daarbij verkregen resultaten (de bevindingen) beter passen bij het hieronder weergegeven Scenario 1 dan bij het alternatieve scenario, Scenario 2.<sup>47</sup>

*Scenario 1: De tape waarmee slachtoffer gekneveld was is afkomstig van de rol tape uit de woning van verdachte.*

*Scenario 2: De tape waarmee slachtoffer gekneveld was is afkomstig van een willekeurige andere rol tape.*

<sup>47</sup> De scenario's zijn geformuleerd op grond van de gegeven informatie over de toedracht.

### **Onderzoek**

De stukken tape waarmee slachtoffer gekneveld was, hebben dezelfde kleur en breedte als de tape op de rol uit de woning van de verdachte. Voorts is geconstateerd dat de lijmlaag van de tijdens de overval gebruikte tape (SVO1.01/SVO1.02) als gevolg van het gebruik verontreinigd is geraakt met vezels van kleding, huidschilfers en enkele andere materialen van vooralsnog onbekende herkomst.

Verder onderzoek is verricht door analyse van de op de tapes aanwezige lijmlaag met *Laser Ablation Inductively Coupled Mass Spectrometry* (afgekort LA-ICP-MS).<sup>48</sup> Dit is een techniek die gebruikt wordt om de concentratie te meten van chemische elementen, zoals ijzer, natrium, zink, goud, koper, enzovoort in allerlei materialen, waaronder papier, glas, verf, en ook tape.

De met LA-ICP-MS verkregen elementprofielen van de tapes die gebruikt zijn bij de overval (SVO1.01/SVO1.02) zijn vergeleken met die van de tape van de in beslag genomen rol (SVO2.01).

### **Resultaten, interpretatie en conclusie**

De chemische profielen van enerzijds SVO1.01/SVO1.02 en anderzijds SVO2.01 komen niet geheel overeen. Dat kan te wijten zijn aan de eerder genoemde verontreiniging. Toch worden bij vergelijking tussen de gebruikte stukken tape met de tape van de rol uit woning van verdachte enkele karakteristieke overeenkomsten in het chemische profiel van de lijmlagen geconstateerd. Die opvallende kenmerken in de chemische profielen zijn niet waargenomen in onze database van lijmlagen van tapes.

### **Conclusie**

Eerder in dit rapport heb ik twee scenario's geformuleerd over de tape die is gebruikt bij de overval. Ik kom nu tot een conclusie. Deze conclusie geeft aan hoeveel *beter of slechter* mijn onderzoeksresultaten passen bij het scenario dat de tape afkomstig is van de rol tape aangetroffen bij de verdachte (Scenario 1) dan bij het scenario dat de tape afkomstig is van een willekeurige andere rol tape (Scenario 2).

Ik formuleer deze conclusie in de vorm van een verhouding. Deze verhouding is de waarschijnlijkheid dat de verkregen resultaten worden aangetroffen als scenario 1 het geval ten opzichte van de waarschijnlijkheid dat de verkregen resultaten worden aangetroffen als scenario 2 het geval is.

---

<sup>48</sup> Indien u meer wilt weten over deze techniek, kunt u onze bijlage 'Materiaal onderzoek met LA-ICP-MS' lezen (nu niet bijgevoegd).



Ik druk deze verhouding uit door een kruisje op onderstaande lijn te plaatsen.

Zeer sterk ten  
gunste van  
Scenario 2

Neutraal

Zeer sterk ten  
gunste van  
Scenario 1



*Ondertekening* Aldus opgemaakt en getekend op de door mij afgelegde  
algemene eed/belofte als vast gerechtelijk deskundige

*Plaats* Den Haag

*Datum* 16 december 2007

# Appendix 2 De cijfers

## STUDIE 1

Tabel A2.1 Feitelijk begrip Studie 1: gemiddeld aantal correct geclassificeerde stellingen

	Gemiddeld aantal goed (Sd)	T-toets tov 4 (sig)	N
Tank iets ws	4.8 (1.06)	13.7 (p < .01)	324
Straat veel ws	4.3 (0.95)	5.2 (p < .01)	330
Bedrijf veel ws'	3.9 (1.15)	-1.5 (ns)	320
Tezamen/3	4.4 (0.71)	8.6 (p < .01)	309

Tabel A2.2 Vermeend begrip Studie 1: vermeend begrip gemeten op 7-puntsschaal  
(1 = 'ik begrijp er niets van'; 7 = 'ik begrijp het helemaal')

	Vermeend begrip (Sd)	% 6 en 7 tezamen	N
Tank (iets ws)	5.8 (1.14)	69	333
Straat (veel ws)	5.9 (1.18)	74	336
Bedrijf (veel ws')	6.2 (0.99)	83	334
Tezamen/3	6.0 (1.00)	76	331

Tabel A2.3 Feitelijk begrip Studie 1: juristen vergeleken met bèta's

	Juristen (Sd)	Bèta's (Sd)	N	T-toets verschil
Tank iets ws	4.92 (1.01)	4.69 (1.09)	323	2.01 (ns)
Straat veel ws	4.25 (0.96)	4.28 (1.95)	329	0.28 (ns)
Bedrijf veel ws'	4.04 (1.09)	3.77 (1.19)	319	2.12 (p = .035)
Tezamen/3	4.43 (0.70)	4.28 (0.72)	308	1.86 (ns)

*Tabel A2.4 Fouten per stelling (%): Tank (iets ws), Straat (veel ws), Bedrijf (veel ws')*

	Tank N = 330	Straat N = 335	Bedrijf N = 330
N ... iets waarschijnlijker stukken afkomstig van rol verdachte dan van	84.8		
N ... veel waarschijnlijker verdachte persoon op beelden dan iemand anders		89.0	
N ... veel waarschijnlijker schoenspoor afkomstig van willekeurige andere schoen dan van verdachte			94.3
N ... iets meer dan 50% kans afkomstig van verdachte	68.5		
N ... veel meer dan 50% kans verdachte persoon beelden		88.1	
N ... veel meer dan 50% kans afkomstig van willekeurige andere			93.0
N ... redelijk bewijs gevonden dat stukken tape afkomstig van verdachte	54.1		
N ... vrij sterk bewijs gevonden dat verdachte op beelden staat		79.4	
N ... vrij sterk bewijs gevonden dat schoenspoor afkomstig van willekeurige andere schoen			83.3
N ... defense fallacy	52.0		
N ... defense fallacy		27.2	
N ... onderzoek niets opgeleverd omdat niet met zekerheid vast te stellen schoenspoor afkomstig van andere schoen dan van verdachte			25.7
J ... uitkomst in enige mate bewijsmateriaal tegen verdachte	19.5		
J ... uitkomst in sterke mate bewijsmateriaal tegen verdachte		27.5	
J ... uitkomst in sterke mate ontlastend bewijsmateriaal voor verdachte			27.4
J ... waarnemingen passen iets beter bij scenario afkomstig van verdachte dan bij scenario	17.3		
J ... waarnemingen passen veel beter bij scenario zelfde persoon dan bij scenario		10.7	
J ... waarnemingen passen veel beter bij scenario afkomstig van willekeurige andere schoen dan bij scenario			8.7
J ... resultaten enigszins belastend voor verdachte	17.3		
J ... resultaten erg belastend voor verdachte		26.5	
J ... resultaten erg ontlastend voor verdachte			32.0
J ... zouden ook van andere rol tape dan van verdachte afkomstig kunnen zijn	6.6		
J ... persoon op beelden zou ook andere persoon dan verdachte kunnen zijn		24.3	
J ... schoenspoor zou ook van schoen van verdachte kunnen zijn			43.9

## STUDIE 2

Tabel A2.5 *Feitelijk begrip Studie 2, tankstation: percentages foute antwoorden per stelling, naar beroepsgroep*

	Rechters (N = 118)	Advocaten (N = 69)	Deskundigen (N = 98)	Verschildtoets $\chi^2$
N... het is iets waarschijnlijker dat stukken afkomstig van rol verdachte dan van ...	92.4	80.9	58.2	p < .01
J... uitkomst in enige mate bewijsmateriaal tegen verdachte ...	53.5	45.5	20.4	p < .01
N... iets meer dan 50% kans afkomstig van verdachte ...	41.4	36.8	36.7	ns
N... defense fallacy ...	36.8	51.5	26.5	p < .01
J... resultaten enigszins belastend voor verdachte ...	35.9	30.9	17.7	p < .05
J... waarnemingen passen iets beter bij scenario afkomstig van verdachte dan bij scenario ...	12.8	15.9	8.2	ns
J... zouden ook van andere rol tape dan die van verdachte afkomstig kunnen zijn ...	4.2	8.8	6.2	ns

Tabel A2.6 *Feitelijk begrip Studie 2, straat: percentages foute antwoorden per stelling, naar beroepsgroep*

	Rechters (N = 118)	Advocaten (N = 69)	Deskundigen (N = 98)	Verschildtoets $\chi^2$
N... het is veel waarschijnlijker dat verdachte de persoon op beelden is dan iemand anders ...	88.0	88.4	63.3	p < .01
J... uitkomst in sterke mate bewijsmateriaal tegen verdachte ...	40.5	28.6	18.4	p < .01
N... veel meer dan 50% kans afkomstig van verdachte ...	70.9	64.7	53.1	p < .05
N... defense fallacy ...	19.7	29.0	12.4	p < .05
J... resultaten erg belastend voor verdachte ...	22.1	20.6	15.3	ns
J... waarnemingen passen veel beter bij scenario afkomstig van verdachte dan bij scenario ...	8.5	13.0	1.0	p < .01
J... persoon op beelden zou ook een andere persoon dan verdachte kunnen zijn ...	13.7	11.8	13.4	ns

**Tabel A2.7 Feitelijk begrip Studie 2: gemiddeld aantal correct geclassificeerde stellingen\***

		Tank, iets ws (Sd)	Straat, veel ws (Sd)	Tezamen/2 (Sd)
1. Rechters (N=118)		4.18 (1.34)	4.31 (1.30)	4.25 (1.17)
2. Advocaten (N=69)		4.22 (1.25)	4.35 (1.10)	4.28 (1.03)
3. Deskundigen (N=98)		5.23 (1.42)	5.21 (1.36)	5.22 (1.27)
Verschillen tussen de beroepsgroepen	<i>Overall</i>	$F_{2,282} = 18.6$ ( $p < .01$ )	$F_{2,282} = 21.7$ ( $p < .01$ )	$F_{2,282} = 21.7$ ( $p < .01$ )
	<i>post-hoc LSD</i>	3 vs 1 & 2 ( $p < .01$ )	3 vs 1 & 2 ( $p < .01$ )	3 vs 1 & 2 ( $p < .01$ )
	<i>Mann-Whitney U</i>	1 vs 3 (3456; $p < .01$ )	1 vs 3 (3727; $p < .01$ )	1 vs 3 (3323; $p < .01$ )
		2 vs 3 (1980; $p < .01$ )	2 vs 3 (2150; $p < .01$ )	2 vs 3 (1876; $p < .01$ )

\* Alle gemiddelden in deze tabel zijn significant groter dan testwaarde 3.5 bij  $p < .01$  (t-toetsen).

**Tabel A2.8 Feitelijk begrip, Studie 2: verschil tussen de twee casus**

	Gemidd. verschil Tank-Straat	Sd	T	df	p
Rechters	-.12	1.30	-.99	117	.32
Advocaten	-.13	1.12	-.96	68	.34
Deskundigen	.02	1.13	.18	97	.86

**Tabel A2.9 Vermeend begrip Studie 2 (tank en straat tezamen)**

		Gemiddeld vermeend begrip (Sd)	% 6 en 7 tezamen	N
1. Rechters		5.28 (1.47)	56	118
2. Advocaten		5.29 (1.50)	59	69
3. Deskundigen		6.15 (0.94)	85	98
Verschillen tussen de beroepsgroepen	<i>Overall</i>	$F_{2,282} = 13.7$ ( $p < 0.01$ )		
	<i>post-hoc LSD</i>	3 vs 1 & 2 ( $p < 0.01$ )		

Tabel A2.10 Vermeend begrip en feitelijk begrip, Studie 2

	Vermeend begrip	Feitelijk begrip	N	Verschil toets	Non-parametrische samenhang (Spearman's rho) Vermeend x Feitelijk
Rechters	laag (<4)	7.53	30	$F_{2,103} = 3.4$ ( $p = 0.04$ )	0.23 ( $p = .01$ )
	goed (4-6)	8.78	50		
	zeer goed (7)	8.88	26		
Advocaten	laag (<4)	7.65	17	$F_{2,58} = 3.8$ ( $p = 0.03$ )	0.23 ( $p = .05$ )
	goed (4-6)	9.23	30		
	zeer goed (7)	8.64	14		
Deskundigen	laag (<4)	10.00	4	$F_{2,77} = 0.30$ ( $p = 0.74$ )	0.09 ( $p = .37$ )
	goed (4-6)	10.24	41		
	zeer goed (7)	10.66	35		

Tabel A2.11 Feitelijk begrip en vermeend begrip Studie 2: verbale en visuele conclusievarianten vergeleken

		Feitelijk begrip			Vermeend begrip		
		Verbaal	Visueel	t-toets	Verbaal	Visueel	t-toets
Rechters	Gemidd.	8.22	8.75	1.22 (ns)	5.02	5.53	1.22 (ns)
	Std.dev.	58	60		1.32	1.58	
	N	2.42	2.24		58	60	
Advocaten	Gemidd.	8.84	8.25	1.18 (ns)	5.30	5.28	1.18 (ns)
	Std.dev.	37	32		1.52	1.50	
	N	1.74	2.37		37	32	
Deskundigen	Gemidd.	10.63	10.26	0.52 (ns)	6.21	6.09	0.52 (ns)
	Std.dev.	51	47		0.95	0.94	
	N	2.45	2.66		51	47	

*Tabel A2.12 Logistische regressies, Studie 2: deskundigen over de prosecutor's fallacy (alle variabelen dichotoom)*

<b>TANK (iets ws)</b>	<b>B</b>	<b>df</b>	<b>p</b>	<b>Exp(B)</b>
Geslacht (man)	0.23	1	0.68	1.26
Leeftijd (>40)	0.45	1	0.43	1.56
Kwant. zelfvertrouwen (>5)	-0.21	1	0.71	0.81
<b>Cursus (ja)</b>	<b>1.76</b>	<b>1</b>	<b>0.04</b>	<b>5.81</b>
Vakliteratuur (ja)	0.41	1	0.46	1.51
Vakbijlage (ja)	0.40	1	0.55	1.49
Tekenbevoegd (ja)	-0.74	1	0.25	0.48
<b>Zelf zo rapporteren (ja)</b>	<b>1.21</b>	<b>1</b>	<b>0.03</b>	<b>3.34</b>
<b>Nagelkerke's <math>R^2 = 0.23</math></b>				
<b>STRAAT (veel ws)</b>	<b>B</b>	<b>df</b>	<b>p</b>	<b>Exp(B)</b>
Geslacht (man)	-0.11	1	0.89	0.89
Leeftijd (>40)	0.48	1	0.42	1.62
Kwant. zelfvertrouwen (>5)	0.86	1	0.15	2.37
<b>Cursus (ja)</b>	<b>2.55</b>	<b>1</b>	<b>0.02</b>	<b>12.86</b>
Vakliteratuur (ja)	0.69	1	0.22	2.00
Vakbijlage (ja)	0.51	1	0.47	1.66
Tekenbevoegd (ja)	-0.34	1	0.60	0.71
<b>Zelf zo rapporteren (ja)</b>	<b>0.55</b>	<b>1</b>	<b>0.33</b>	<b>1.74</b>
<b>Nagelkerke's <math>R^2 = 0.27</math></b>				





## Appendix 3 Gehele vragenlijst, verbale variant, voor juristen

### INTRODUCTIE

Het Nederlands Studiecentrum Criminaliteit en Rechtshandhaving (NSCR) verricht, in samenwerking met het Nederlands Forensisch Instituut (NFI), onderzoek naar uiteenlopende aspecten van deskundigenrapportages. In het kader daarvan vragen wij uw medewerking.

Op de volgende pagina's treft u een tweetal rapporten aan van deskundigen die een onderzoek hebben verricht naar aanleiding van een misdrijf.

Over ieder van die rapporten willen wij u enkele vragen stellen. Wilt u a.u.b. nadat u het eerste rapport heeft gelezen eerst de daarbij behorende vragen beantwoorden en pas daarna doorgaan naar het volgende rapport?

De gegevens worden anoniem verwerkt.

Het invullen van de vragenlijst duurt ongeveer 20 minuten.

**Wij zijn u bij voorbaat zeer dankbaar voor uw medewerking!**

## DESKUNDIGENRAPPORT

### Beroving op straat

*Onderwerp* Gezichtsvergelijking naar aanleiding van een beroving onder bedreiging van een mes bij een pinautomaat te Den Haag

#### Onderzoeksmateriaal

*Ontvangen via* Technische Recherche Hollands Midden  
*Datum ontvangst* 28 december 2007

Code	Omschrijving
TR_A1	Cd met daarop digitaal beeldmateriaal; inhoud als volgend hieronder
SX1_B1	Foto frontaal aangezicht verdachte
SX1_B2	Foto profiel rechterzijde gezicht verdachte
SX1_B3	Foto profiel linkerzijde gezicht verdachte
SX1_C1	Stilstaand beeld van beveiligingscamera met onderschrift 'tijdcode 21:32:03'
SX1_C2	Stilstaand beeld van beveiligingscamera met onderschrift 'tijdcode 21:32:07'
SX1_C3	Stilstaand beeld van beveiligingscamera met onderschrift 'tijdcode 21:32:22'
SX1_C4	Stilstaand beeld van beveiligingscamera met onderschrift 'tijdcode 21:32:27'
SX1_C5	Stilstaand beeld van beveiligingscamera met onderschrift 'tijdcode 21:32:39'
SX1_C6	stilstaand beeld van beveiligingscamera met onderschrift 'tijdcode 21:33:01'

#### *Toelichting op onderzoeksmateriaal*

De beelden SX1\_C1 t/m SX1\_C6 zijn afkomstig van de beveiligingscamera bij de pinautomaat. Verbalisant van de TR lichtte toe dat op ieder van die beelden de dader van de beroving voorkomt. Voor zover er twee personen te zien zijn, betreft het dader en slachtoffer (dit is het geval bij SX1\_C2, SX1\_C4, SX1\_C5). Dader is in deze beelden telkens te herkennen aan de baseballpet, terwijl slachtoffer geen hoofddekseel draagt.

#### Aanleiding onderzoek

De volgende informatie is geleverd door de verbalisant.

Een jongeman die 's avonds op straat bij een pinautomaat zojuist € 100 heeft opgevraagd, wordt door een onbekende plotseling onder bedreiging van een mes dat bedrag afhandig gemaakt en gedwongen om direct nog eens € 500 te pinnen en af te geven. Naar aanleiding van een anonieme tip houdt de politie diezelfde avond een verdachte aan in een café. De technische recherche onderzoekt de volgende ochtend de opnames van de beveiligingscamera bij de pinautomaat en kopieert daarvan een zestal stilstaande beelden waarop de dader van de beroving zichtbaar is. Van verdachte worden foto's gemaakt op het politiebureau.

**Vraagstelling**

*Is verdachte op de foto's SX1\_B1, SX1\_B2 en SX1\_B3 dezelfde persoon als de dader van de beroving op de beelden SX1\_C1 t/m SX1\_C6?*

Na een eerste screening van het aangeleverde beeldmateriaal op de cd (TR\_A1) werd duidelijk dat belichting, contrast en scherpte van de SX1\_C2 t/m SX1\_C6 van dusdanig slechte kwaliteit zijn, dat deze beelden ongeschikt zijn voor gezichtsvergelijking. SX1\_C1 is van matige kwaliteit, maar voldoende om in aanmerking te komen voor gezichtsvergelijking. Het betreft een beeld waarop de dader frontaal te zien is.

De vraagstelling is geïnterpreteerd als het verzoek om een vergelijkend onderzoek te verrichten en daarbij na te gaan of de bevindingen beter passen bij de hieronder weergegeven hypothese H1 dan bij de (alternatieve) hypothese H2.<sup>49</sup>

*Hypothese 1: De dader van de beroving zichtbaar op SX1\_C1 is dezelfde persoon als verdachte afgebeeld op foto SX1\_B1.*

*Hypothese 2: De dader van de beroving zichtbaar op SX1\_C1 is niet dezelfde persoon als verdachte afgebeeld op foto SX1\_B1.*

**Onderzoek**

De gebruikte methode voor persoonsidentificatie aan de hand van gezichtsvergelijking op beeldmateriaal is gebaseerd op het vergelijken van morfologisch-antropologische kenmerken. Het aantal en de mate van gelijkheid van de gevonden kenmerken zijn de basis voor de gerapporteerde mate van steun voor de hypothesen dat de te vergelijken afbeeldingen wel of niet dezelfde persoon betreffen. Met behulp van software is het beeldmateriaal eerst zo veel mogelijk geoptimaliseerd door aanpassing van contrast, helderheid en inzoomen op details.

**Resultaten, interpretatie en conclusie**

Vergelijking van de gezichtskenmerken van de personen op foto SX1\_C1 en SX1\_B1 geeft een veelheid aan overeenkomsten in morfologische en antropologische kenmerken. Uit wetenschappelijk onderzoek is mij bekend dat deze combinatie van gelaatskenmerken in de Nederlandse bevolking zeldzaam is. Ik concludeer nu het volgende:

49 De hypothesen zijn geformuleerd op grond van de gegeven informatie over de toedracht.

De bevindingen van de hier gerapporteerde gezichtsvergelijking aan de hand van het geselecteerde beeldmateriaal zijn *veel waarschijnlijker* wanneer de daarop afgebeelde persoon dezelfde persoon is (hypothese 1) dan wanneer het een andere persoon betreft (hypothese 2).<sup>50</sup>

*Ondertekening* Aldus opgemaakt en getekend op de door mij afgelegde algemene eed/belofte als vast gerechtelijk deskundige

*Plaats* Den Haag

*Datum* 11 januari 2008

Wilt u nu eerst enkele algemene vragen over dit rapport beantwoorden? Indien u meer ruimte nodig heeft, dan kunt u de achterzijde van het blad gebruiken.

1. Kunt u hieronder aangeven hoe goed of slecht u de conclusie van dit rapport begrijpt? (omcirkel het getal dat het beste bij uw mening past)

ik begrijp er  
niets van

ik begrijp het  
volledig

1

2

3

4

5

6

7

2. Als u de conclusie niet helemaal goed begrijpt, waar ligt dat dan volgens u aan?

.....  
.....  
.....

3. Kunt u de conclusie van dit rapport in uw eigen woorden opschrijven?

.....  
.....  
.....

<sup>50</sup> *Toelichting op de conclusie*

Voor het vergelijkend onderzoek wordt gebruikgemaakt van de volgende conclusiereeks waarmee een uitspraak wordt gedaan over de waarschijnlijkheid van de bevindingen, in het licht van de gestelde hypothesen. De gebruikte reeks is:

De bevindingen van het onderzoek zijn ...

- ongeveer even waarschijnlijk ... [als]
- iets waarschijnlijker ...
- waarschijnlijker ...
- veel waarschijnlijker ...
- zeer veel waarschijnlijker ...

... wanneer hypothese 1 juist is, dan wanneer hypothese 2 juist is.

Hieronder volgt een aantal stellingen over dit rapport. Geef a.u.b. bij ieder van de stellingen aan of dit een correcte interpretatie van de conclusie in het rapport is.

		Is dit een correcte interpretatie van de conclusie?		
		Ja	Nee	Weet niet
1	Er is veel meer dan 50% kans dat de verdachte de persoon op de camerabeelden is.			
2	Het is veel waarschijnlijker dat de verdachte de persoon op de camerabeelden is dan dat het iemand anders is die op de camerabeelden staat.			
3	De uitkomst van dit onderzoek vormt in sterke mate bewijsmateriaal tegen de verdachte.			
4	Er is aangetoond dat de verdachte schuldig is.			
5	De resultaten van dit onderzoek zijn erg belastend voor de verdachte.			
6	Er zijn treffende overeenkomsten gevonden tussen het gelaat van de verdachte en het gelaat van de persoon op de camerabeelden. Het valt echter niet uit te sluiten dat een aantal andere mensen, net als de verdachte, lijkt op de persoon op de camerabeelden. De verdachte maakt dus deel uit van een groep mensen die allemaal op de beelden zouden kunnen staan. De kans is dus erg klein dat juist de verdachte op de camerabeelden staat.			
7	De waarnemingen passen veel beter bij het scenario dat de verdachte dezelfde persoon is als de persoon op de camerabeelden, dan bij het scenario dat de persoon op de camerabeelden iemand anders is.			
8	De persoon op de camerabeelden zou ook een andere persoon dan de verdachte kunnen zijn.			

## DESKUNDIGENRAPPORT

### Beroving tankstation

*Onderwerp* Chemische profilering van tape in relatie tot gewapende overval tankstation

#### **Onderzoeksmateriaal**

Ontvangen via KLPD  
Datum ontvangst 12 december 2007

Code	Omschrijving
SVO1.01	Stuk gebruikte blauwe tape gebruikt tijdens overval, lengte ca. 83 cm
SVO1.02	Stuk gebruikte blauwe tape gebruikt tijdens overval, lengte ca. 98 cm
SVO2.01	Rol blauwe tape uit woning verdachte

#### **Aanleiding onderzoek**

De volgende informatie is geleverd door de verbalisant.

Bij een gewapende overval op een tankstation langs de N13 heeft de gemaskeerde overvaller tape gebruikt om de in het tankstation aanwezige pompbediende te knevelen. Naar aanleiding van een anonieme tip is een dag na de overval een verdachte in zijn woning aangehouden. Bij zoeking in de woning vond de politie op het aanrecht een rol blauwe tape die in beslag is genomen voor forensisch onderzoek.

#### **Vraagstelling**

*Onderzoek de mate van overeenkomst tussen de tapes van SVO1.01/ SVO1.02 met de rol tape van SVO2.01.*

De vraagstelling is geïnterpreteerd als het verzoek om een vergelijkend onderzoek te verrichten en na te gaan of de daarbij verkregen resultaten (de bevindingen) beter passen bij de hieronder weergegeven hypothese H1 dan bij de (alternatieve) hypothese H2.<sup>51</sup> De hypothesen luiden:

Hypothese 1: *De tape waarmee slachtoffer gekneveld was is afkomstig van de rol tape uit de woning van verdachte.*

Hypothese 2: *De tape waarmee slachtoffer gekneveld was is afkomstig van een willekeurige andere rol tape.*

<sup>51</sup> De hypothesen zijn geformuleerd op grond van de gegeven informatie over de toedracht.

### Onderzoek

De stukken tape waarmee slachtoffer gekneveld was, hebben dezelfde kleur en breedte als de tape op de rol uit de woning van de verdachte. Voorts is geconstateerd dat de lijmlaag van de tijdens de overval gebruikte tape (SVO1.01/SVO1.02) als gevolg van het gebruik verontreinigd is geraakt met vezels van kleding, huidschilfers en enkele andere materialen van vooralsnog onbekende herkomst.

Verder onderzoek is verricht door analyse van de op de tapes aanwezige lijmlaag met *Laser Ablation Inductively Coupled Mass Spectrometry* (afgekort LA-ICP-MS).<sup>52</sup> Dit is een techniek die gebruikt wordt om de concentratie te meten van chemische elementen, zoals ijzer, natrium, zink, goud, koper, en zovoort in allerlei materialen, waaronder papier, glas, verf, en ook tape.

De met LA-ICP-MS verkregen elementprofielen van de tapes die gebruikt zijn bij de overval (SVO1.01/SVO1.02) zijn vergeleken met die van de tape van de in beslag genomen rol (SVO2.01).

### Resultaten, interpretatie en conclusie

De chemische profielen van enerzijds SVO1.01/SVO1.02 en anderzijds SVO2.01 komen niet geheel overeen. Dat kan te wijten zijn aan de eerder genoemde verontreiniging. Toch worden bij vergelijking tussen de gebruikte stukken tape met de tape van de rol uit woning van verdachte enkele karakteristieke overeenkomsten in het chemische profiel van de lijmlagen geconstateerd. Die opvallende kenmerken in de chemische profielen zijn niet waargenomen in onze database van lijmlagen van tapes.

Ik kom nu tot de volgende conclusie:

De bevindingen van het vergelijkend onderzoek zijn *iets waarschijnlijker* wanneer de tape waarmee slachtoffer gekneveld was afkomstig is van de rol tape uit de woning van verdachte (hypothese 1) dan wanneer de tape waarmee slachtoffer gekneveld was afkomstig is van een willekeurige andere rol tape (hypothese 2).<sup>53</sup>

52 Indien u meer wilt weten over deze techniek, kunt u onze bijlage 'Materiaal onderzoek met LA-ICP-MS' lezen (nu niet bijgevoegd).

53 *Toelichting op de conclusie*

Voor het vergelijkend onderzoek wordt gebruikgemaakt van de volgende conclusiereeks waarmee een uitspraak wordt gedaan over de waarschijnlijkheid van de bevindingen, in het licht van de gestelde hypothesen. De gebruikte reeks is:

De bevindingen van het onderzoek zijn ...

- ongeveer even waarschijnlijk ... [als]
- iets waarschijnlijker ...
- waarschijnlijker ...
- veel waarschijnlijker ...
- zeer veel waarschijnlijker ...

... wanneer hypothese 1 juist is, dan wanneer hypothese 2 juist is.



*Ondertekening* Aldus opgemaakt en getekend op de door mij afgelegde  
algemene eed/belofte als vast gerechtelijk deskundige  
*Plaats* Den Haag  
*Datum* 16 december 2007

Wilt u nu eerst enkele algemene vragen over dit rapport beantwoorden?  
Indien u meer ruimte nodig heeft, dan kunt u de achterzijde van het blad  
gebruiken.

1. Kunt u hieronder aangeven hoe goed of slecht u de conclusie van dit rap-  
port begrijpt? (omcirkel het getal dat het beste bij uw mening past)
- ik begrijp er  
niets van

ik begrijp het  
volledig
- 1

2

3

4

5

6

7
2. Als u de conclusie niet helemaal goed begrijpt, waar ligt dat dan volgens  
u aan?
- 
3. Kunt u de conclusie van dit rapport in uw eigen woorden opschrijven?
- 

Hieronder volgt een aantal stellingen over dit rapport. Geef a.u.b. bij ieder  
van de stellingen aan of dit een correcte interpretatie van de conclusie in het  
rapport is.

		Is dit een correcte interpretatie van de conclusie?		
		Ja	Nee	Weet niet
1	Er is iets meer dan 50% kans dat de stukken tape afkomstig zijn van de rol van de verdachte.			
2	Het is iets waarschijnlijker dat de stukken tape afkomstig zijn van de rol van de verdachte dan van een andere rol.			
3	De uitkomst van dit onderzoek vormt in enige mate bewijsmateriaal tegen de verdachte.			
4	Er is aangetoond dat de verdachte schuldig is.			
5	De resultaten van dit onderzoek zijn enigszins belastend voor de verdachte.			

6	Er zijn enkele overeenkomsten gevonden tussen de stukken gebruikte tape en de tape op de rol van de verdachte. Het valt echter niet uit te sluiten dat er een aantal rollen tape in omloop is die even sterke overeenkomsten vertoont met de gevonden stukken tape als de rol van de verdachte. De rol tape van de verdachte maakt dus deel uit van een groep rollen die allemaal op elkaar lijken. De kans is dus erg klein dat de gebruikte tape juist van de rol tape van verdachte komt.			
7	De waarnemingen passen iets beter bij het scenario dat de stukken tape afkomstig zijn van de rol van de verdachte dan bij het scenario dat ze van een andere rol afkomstig zijn.			
8	De stukken tape zouden ook van een andere rol tape dan die van de verdachte afkomstig kunnen zijn.			

Wij willen u nog enkele algemene vragen stellen over de wijze waarop deskundigen wel eens conclusies formuleren en hoe dat op u overkomt.

1. Deskundigen maken in hun onderzoek en rapportage vaak gebruik van hypothesen (Hypothese 1 en Hypothese 2).  
Kunt u kort aangeven wat u verstaat onder ‘een hypothese’?  
.....  
.....  
.....
2. Soms wordt geprobeerd de conclusie van de deskundige in een getal uit te drukken. In de tabel staan hierbij verschillende mogelijkheden.  
Wilt u de tabel zorgvuldig bekijken en vervolgens daaronder aangeven welke van de mogelijkheden volgens u voor de deskundige de beste manier is om de conclusie in een getal uit te drukken?

Huidige weergave	Weergave met getallen			
De bevindingen van het onderzoek zijn (...) wanneer hypothese 1 waar is dan wanneer hypothese 2 waar is.	De bevindingen van het onderzoek zijn (...) maal zo waarschijnlijk onder hypothese 1 dan onder hypothese 2.			
	Mogelijkheid 1	Mogelijkheid 2	Mogelijkheid 3	Liever mijn eigen weergave, en wel ...
‘ongeveer even waarschijnlijk’	1	1	1	...
‘iets waarschijnlijker’	2	2	10	...
‘waarschijnlijker’	3	4	100	...
‘veel waarschijnlijker’	4	8	1000	...
‘zeer veel waarschijnlijker’	5	16	10000	...

De beste manier is:

- o Mogelijkheid 1
- o Mogelijkheid 2
- o Mogelijkheid 3
- o Mijn eigen weergave
- o Geen van de mogelijkheden

Wij verzoeken u ten slotte de volgende algemene vragen te beantwoorden.

		Ja	Nee
1.	Heeft u in de afgelopen jaren een cursus gevolgd over onderwerpen zoals de wijze waarop deskundigen hun conclusies formuleren en de interpretatie daarvan in de rechtszaal?		
2.	Heeft u in de afgelopen jaren symposia of congressen bezocht over onderwerpen zoals deze?		
3.	Leest u met enige regelmaat vakliteratuur of wetenschappelijke literatuur over deze onderwerpen?		
4.	Heeft u de vakbijlage van het NFI over dit onderwerp gelezen ('De reeks waarschijnlijkheidstermen van het NFI en het Bayesiaanse model voor de interpretatie van bewijs')?		

5. Als deskundigenrapporten niet goed begrepen worden, is dat een probleem dat de deskundige dient op te lossen.

Helemaal oneens

1234567

Helemaal eens

6. De professionele procesdeelnemers dienen zich bij te scholen om technisch-forensische rapporten goed te kunnen begrijpen.

Helemaal oneens

1234567

Helemaal eens

7. Het grote verschil in vakspecifieke kennis en referentiekader tussen technisch deskundige en jurist zorgt ervoor dat iedere dialoog tussen beide altijd gekenmerkt wordt door kloven en misverstanden.

Helemaal oneens

1234567

Helemaal eens

8. Hoe lang bent u in het strafrecht werkzaam?

.....

9. Hoe vaak ziet u per jaar een technisch-forensische deskundigenrapportage?  
 ..... keer per jaar (bij benadering)
10. In technisch-forensische rapporten worden conclusies in toenemende mate geformuleerd in termen van de relatieve waarschijnlijkheid van de bevindingen onder twee hypothesen of scenario's.  
 Bent u deze wijze van concluderen in concrete strafzaken al tegengekomen?  
 o Ja  
 o Nee
11. Wat is uw mening over deze wijze van concluderen?  
 .....  
 .....  
 .....
12. Bent u van mening dat er binnen uw beroepsgroep voldoende specifieke kennis aanwezig is om goed te begrijpen wat deskundigen in hun rapportages tot uitdrukking trachten te brengen?  
 o Ja, zonder meer  
 o Ja, maar het houdt niet over  
 o Nee, men schiet hierin enigszins tekort  
 o Nee, bij lange na niet
13. Hoe vertrouwd voelt u zichzelf als het gaat om het lezen en begrijpen van cijfermateriaal en statistische gegevens? (Omcirkel het getal dat het beste bij uw mening past)
- |   |   |   |   |   |   |  |
|---|---|---|---|---|---|--|
| ik voel me <b>helemaal niet</b><br>vertrouwd in het omgaan<br>met cijfers |   |   |   |   |   | ik voel me <b>heel erg</b><br>vertrouwd in het omgaan<br>met cijfers |
| 1   | 2 | 3 | 4 | 5 | 6 | 7  |
14. Wat is uw geslacht?  
 o Man  
 o Vrouw
15. Wat is uw leeftijd?  
 ..... jaar

Nogmaals veel dank voor uw medewerking!



## Over de auteurs

*Prof. dr. Henk Elffers* studeerde mathematische statistiek en waarschijnlijkheidsrekening aan de Universiteit van Amsterdam en promoveerde op een studie naar de psychologie van de belastingontduiking aan de Erasmus Universiteit Rotterdam. Hij is deeltijd hoogleraar empirische bestudering van de strafrechtpleging aan de Vrije Universiteit Amsterdam en senior onderzoeker bij het Nederlands Studiecentrum Criminaliteit en Rechtshandhaving. Zijn onderzoeksinteresse gaat uit naar rationele-keuzetheorie van de regelovertreiding, ruimtelijke aspecten van criminaliteit, simulatiestudies in de criminologie, en de verhouding burger-strafrechter.

*Dr. Jan W. de Keijser* is senior onderzoeker bij het Nederlands Studiecentrum Criminaliteit en Rechtshandhaving (NSCR) te Amsterdam. Hij studeerde politicologie aan de Universiteit Leiden en promoveerde in 2000 in de sociale wetenschappen op een proefschrift over straftoemeting door Nederlandse rechters en de rol van strafdoelen daarin. Zijn onderzoeksbelangstelling gaat onder andere uit naar de wisselwerking tussen rechter en samenleving, straftoemeting, bewijs en de bewijsbeslissing, en publieke opinie over de strafrechtspleging.

*Drs. Roos Marijn Kok* is thans werkzaam als gezinsvoogd bij Bureau Jeugdzorg in Amsterdam. Daarvoor werkte zij als junior onderzoeker bij het Nederlands Studiecentrum Criminaliteit en Rechtshandhaving mee aan het onderzoek waarvan in deze uitgave verslag wordt gedaan. Zij studeerde psychologie aan de Universiteit Leiden.

*Dr. Marjan J. Sjerps* studeerde wiskunde aan de Katholieke Universiteit Nijmegen (thans Radboud Universiteit). Zij studeerde in 1988 af in de richting statistiek en operations research. Daarna verrichtte zij promotieonderzoek naar evolutionair stabiele strategieën in de gedragsecologie aan de Universiteit Leiden. In 1994 verscheen haar proefschrift over dit onderwerp. Sinds 1993 is zij werkzaam als statisticus op het Nederlands Forensisch Instituut, momenteel als teamleidster van de statistiekgroep. Haar onderzoeksbelangstelling gaat uit naar de toepassing van statistiek en kansrekening op forensisch bewijs, en het gebruik daarvan in de strafrechtsketen.